34th Annual Conference of the Military Testing Association

PROCEEDINGS

Volume II

Prepared and Coordinated by

Navy Personnel Research and Development Center
San Diego, California

26 - 29 October 1992

34th Annual Conference of the Military Testing Association

Hosted by the

Navy Personnel Research and Development Center
53335 Ryne Road
San Diego, California 92152-7250

26 - 29 October 1992

Conference Committee

Commanding Officer, NPRDC	*CAPT Thomas Finley*
Chairs, MTA-92 Conference	*Drew Sands* *John Ellis*
Chair, Program Committee	*John Ellis*
Chair, Registration Committee	*Margie Sands*
Chair, Hospitality Committee	*John Ellis*
Chair, Publication Committee	*Drew Sands* *John Ellis*

Acknowledgments

The members of the MTA-1992 Conference Committee wish to express their appreciation for the expertise and dedication of the following individuals. Each person had an important role in the success of the 34th Annual Conference of the Military Testing Association.

Audiovisual Support

Jim Julius

Database

Margie Sands
Anthy Dunlap
Christina Reese

Finance

John Ellis
Drew Sands

Hospitality

John Ellis
Kathy Ellis

Program Organization

Margie Sands

Program Production

John Ellis
Margie Sands
Ruth Ireland
Mely Leano
Marci Barrineau

Proceedings Publication

Drew Sands
John Ellis
Marci Barrineau
Carmen Fendelman

Registration

Margie Sands
Anthy Dunlap
Christina Reese
Peggy Laone

Session Chairs

Herb Baker
Mike Cowen
Ronna Dillon
John Ellis
John Folchi
Paul Foley
Dennis Gettman
Alice Gerb
Janet Held
Rebecca Hetter
Jerry Laabs
Gerald Larson
Reynaldo Monzon
Robert Morrison
Randolph Park
Shelley Perry
Josephine Randel
Malcolm Ree
Dave Robertson
Carol Robinson
Ellie Robinson
Hendrick Ruck
Michael Rumsey
Drew Sands
Dan Segall
Steve Sellman
George Seymour
Wallace Sinaiko
Mannie Somer
Friedrich Steege
Hervey Stern
Walt Thode
Tom Trent
Wolfgang Weber
John Welsh
Lauress Wise
Martin Wiskoff

Preceding Page Blank

Foreword

The Proceedings of the 34th Annual Conference of the Military Testing Association document the presentations given during the Conference paper and panel sessions. The papers present a broad range of topics by contributors from the military, private industry, and the educational communities. It should be noted that the papers reflect the opinions of the authors and do not necessarily reflect the official policy of any military service, institution, or government.

Contents

Session 24

Self Assessment and Human Learning Skills	*G. F. Fullerton* *D. P. Hunt*	497
How Personnel Testing is Adapting to a Changing Air Force	*T. S. Ziebell*	503
Problems of Exam Compromise	*L. J. Paquette* *R. E. Doucette*	509
The Effects of UH-1 Experience on UH-60 Simulator Performance: A Preliminary Study	*C. A. Salter* *J. S. Crowley* *J. A. Caldwell* *R. L. Smith*	514
Auditory Guidance in Officer Level Training	*R. O. Waldkoetter* *P. L. Vandivier*	520
Components and Metacomponents of Intelligence Among Navy and Air Force Personnel	*R. F. Dillon*	526

Session 25

Measuring Martial Attitudes: The Military Ethos Scale (MES) in Retrospect	*S. B. Flemming*	532
Stress Effects of Army Downsizing on Family Members	*M. E. Freville*	538
The Socio-Economic Benefits of Home-Basing of Army Units	*H. Lakhani*	547
Special Forces Prior Service Program	*E. J. Brady*	553
Demographics and Aptitude	*M. J. Ree* *J. A. Earles*	559
Assessing the Coachability of Project A Spatial Tests: Developing Strategies	*D. R. Palmer* *H. H. Busciglio*	564
Assessing the Coachability of Project A Spatial Tests: Empirical Results	*H. H. Busciglio*	570

Session 26

A Principal Components Analysis of 59 Variables Descriptive of Uncovered Spies	*L. A. Stone*	576
Turnover Intentions During Naval Officer Training	*J. M. Stouffer*	582
Vertical Slice Battalion-Level Assessment of Future Command and Control Systems	*C. W. Lickteig* *S. G. Williams* *D. L. Smart*	588
Longitudinal Research on Officer Careers (LROC): Impact of Change on Attitudes and Careers	B. C. Harris	594

Session 27

The Validity of Two Methods of Testing Flexibility	*R. F. Dillon* *R. J. Greer*	600
The Validity of Information-Processing Parameters of Inductive Reasoning	*R. F Dillon* *C. A. Harris*	605
Further Construct Validation of Cognitive Flexibility	*R. F. Dillon* *T. S. Brannan*	610
Emotional Control, Inductive Reasoning, Academic Performance	*R. Webb* *R. F. Dillon*	616
A Template for Evaluating Studies Presenting Potential Predictors of Effectiveness	*W. F. Kieckhaefer*	620

Session 28

Espionage by U.S. Citizens: 1945-1990	*M. F. Wiskoff* *S. Wood*	624
Espionage and Betrayal of Trust	*J. P. Parker* *M. F. Wiskoff*	630
Assessing the Espionage Vulnerability of Positions	*K. S. Crawford* *M. J. Bosshardt*	636
Getting Serious About Security Awareness: Better Briefings, Videos, and Posters Alone Won't Do It	*J. A. Riedel* *J. P. Parker*	642

X

Session 29

The Changing German Armed Forces in 1992	*F. W. Steege*	648
Meeting the German Armed Forces Requirements for Conscripts: A Current Challenge	*H. J. Ebenrett*	654
Equal Opportunity for East and West German Volunteers: An Unresolved Issue	*G. W. Rodel*	660
Aptitude and Motivation of Officer Applicants Selected and Rejected	*A. H. Melter*	665
Computer-Assisted Programs in the Training of Leadership Behavior	*E. W. Bucher*	671

Session 30

Building a Joint-Service Classification Research Roadmap: Current Classification Procedures	*T. L. Russell* *D. J. Knapp* *J. P. Campbell*	676
Building a Joint-Service Classification Research Roadmap: Predictor-Related Research Needs	*D. H. Reynolds* *T. L. Russell* *J. P. Campbell*	682
Building a Joint-Service Classification Research Roadmap: Criterion-Related Research Needs	*D. J. Knapp* *J. P. Campbell*	688
Building a Joint-Service Classification Research Roadmap: Classification Research Objectives	*J. P. Campbell* *T. L. Russell* *D. J. Knapp* *D. H. Reynolds*	694

Session 31

The ASVAB Test Content Specifications	*J. A. Harris*	700
Test Accuracy Specification	*L. L. Wise*	706
The ASVAB Item Development	*J. G. Guzaitis* *G. Glick*	712
The ASVAB Item Tryout Study	*J. R. Welsh*	718

Session 32

Analysis Levels and Inter-Level Linkages: Moderating Change and Intervention Effects	*W. R. Bennett, Jr.* *H. W. Ruck* *A. I. Huffcutt*	724
Multilevel Occupational Analysis: Hierarchies of Tasks, Modules, Jobs, and Specialties	*J. L. Mitchell* *W. J. Phalen* *D. K. Hand*	729
An Organizational Analysis Simulation Technology	*D. S. Vaughan* *R. M. Yadrick*	735

Session 33

ODMARS: An Occupational Analysis Productivity Enhancement Tool	*J. S. Price*	741
Differences in Content Validity Ratings Across Air Force Career Fields and Grade Levels (SENDING SINGLE SIDED COPY)	*T. K. Mayhill* *J. C. Harris* *D. W. Schuette* *P. P. Stanley II*	745
Rationale for Cognitive Task Analysis of Tactical Decision Making	*J. A. Modrick*	751
Training Emphasis Based on Post-Operation Desert shield/Storm Occupational Surveys	*L. A. Goldman*	757
The Role of Testing in Advanced Industrial Management	*L. S. Buck*	763
Naval Officer Computer Utilization	*K. A. Doyle*	768
Using Event History Analysis to Study Task Data: An Update	*S. D. Stephenson* *J. A. Stephenson*	774
Navy Occupational Analysis in a Total Force Environment	*D. L. Moore*	780

Session 34

Attrition Management Tool: Navy Recruit Psychological Screening	*I. Idar* *J. A. Scaramozzino*	786
Characteristics of VSI/SSB Takers: Issues for Testing	*K. A. Martell* *J. M. Arabian*	792

Session 35

Commander Survivability on the NTC Battlefield	*R. F. Holz*	798
Using Simulation to Support Testing: Implications of a HARDMAN III Application	*L. Allender* *D. M. McAnulty* *C. Bierbaum*	804
On Minimizing Fratricide Risks	*G. L. Neal*	810
Occupational Survey Report	*H. Huguley III*	816

Session 36

Utility of Occupational Surveys for Assessing Enlisted Job Performance	*M. R. Miller* *J. Skinner* *D. L. Harville*	821
Enlisted Job Classification Technology: Initial Development	*W. E. Driskill* *J. J. Weissmuller* *M. J. Dittmar*	827
Development of a Career Field Management Plan for the Electronic Computer and Switching Systems Specialty	*T. Coccia* *J. L. Mitchell* *J. R. Knight* *R. C. Shrum*	831

Session 37

Measurement Equivalence and Score Equating for Experimental and Operational Basic Attributes Test	*T. R. Carreta* *M. J. Ree*	837
Introduction to Back Propagation: An Artificial Neural Network Paradigm	*W. A. Sands*	843
The Ultimate Person-Job Match: A Key to Future Worker Productivity	*J. H. Ward, Jr.* *D. S. Vaughan* *J. L. Mitchell* *W. E Driskill* *H. W. Ruck*	849

Session 38

An Integrated Training Systems Model for Modular Control Equipment	*E. G. Gibson*	855
An Initial Evaluation of the use of Captioned Television to Improve the Vocabulary and Reading Comprehension of Navy Sailors	*G. R. Griffin* *J. C. Dumestre*	861
Students' Reception and Discovery Methods Affected by Their Information Processing Behavior	*W. Rodgers*	867
Replacing Paper Users Manuals with On-Line Help	*D. B. Moskow*	872

Session 39

Navy-wide Personnel Survey (NPS) 1991: Everyone Has an Opinion	*M. A. Quenette*	877
Empathizing with the Survey Respondent	*G. L. Wilcove*	883
Let's Hear from the Reserves: Results of the 1991 Naval Reserve Survey	*H. G. Baker*	888

Session 40

The Factor Structure of Cognitive, Spatial, Perceptual, and Psychomotor Tests	*N. G. Peterson* *R. L. Rosse* *T. L. Russell*	894
The Factor Structure of a Spatial Test Battery	*T. L. Russell* *L. Humphreys* *R. L. Rosse* *N. G. Peterson*	900
General Aptitude and Specific Language Learning Aptitude	*J. W. Thain*	906

Session 41

What's Past is Prologue	*M. J. Ree*	912
Selection and Classification for the Career Force	*M. G. Rumsey*	918
FINIS CORONAT OPUS	*F. L. Vicino*	923

Session 42

Hispanic Underrepresentation in Navy's Blue-Collar Civilian Labor Force: Organizational Perspectives	*P. Rosenfeld* *J. E. Edwards* *M. D. Thomas* *S. Booth-Kewley*	929
Equal Opportunity Perceptions of Filipinos in the U. S. Navy	*D. E. Silva-Jalonen* *P. Rosenfeld*	933
Supervisor's Gender and Race Affect Navy Black Females Equal Opportunity Perceptions	*C. E. Newell* *P. Rosenfeld* *A. L. Culbertson*	938

Author Index I-1

Self Assessment and Human Learning Skills

By
George F. Fullerton and Darwin P. Hunt, Ph.D.
Human Performance Enhancement, Inc.

ABSTRACT

Human self assessment related to human learning, testing and performance has been the subject of over a decade of research by Dr. Darwin P. Hunt. The results of this research program clearly identified self assessment as an important factor in the processes by which one acquires and uses knowledge.

It was recognized, early on, that in today's environment of high technology, that the speed and accuracy of a person's responses, when interfacing with highly complex systems and machinery, is key. The skill of the human operator can make the difference between success or failure.

A unique system of testing that provides a quantitative and qualitative measurement of a person's usable knowledge is desribed in detail. The testing system identifies accurately the well informed, uninformed, and the misinformed. The system, known as the Self Assessment Computer Assisted Test (SACAT) and SACAT-San (a paper-and-pencil version), developed by Human Performance Enhancement, Inc. provides an improved basis for the measurement of the effectiveness of the training program.

SELF ASSESSMENT TECHNOLOGY

SACAT and SACAT Scan were developed using as it's basis the standard and commonly used multiple choice test. The multiple choice test simply requires the person to select an answer from among a list of alternative answers.

The multiple choice test has many advantages which include ease and objectivity of scoring, the ability to measure simple and complex objectives in most subject areas at most levels of knowledge, the ability to sample domains of knowledge, the ability to determine whether specified objectives regarding learning have been reached, reliability and economy/efficiency. For those reasons it is expected that the multiple choice tests will be used more and more often.

However, the usual multiple choice test has some limitations and disadvantages, which are unacceptable in some cases, such as (1) it emphasizes the recognition of the correct answer rather than recall and (2) it provides a single measure - a percent correct score - of the person's knowledge on the topics of the test. The emphasis on recognition encourages the test taker to guess at answers with no particular conviction that their answers are correct. Such unsure responses do not represent the kind of judgement that govern everyday decisions and actions. Also guessing allows the test taker to get some answers correct by chance, which inflates the percent correct score.

The knowledge of a person has many more characteristics than is represented by the percent correct score. The Self Assessment (SA) Test incorporates the idea that there are various degrees to which a person knows something. By allowing test takers to indicate their doubt or certainty about correctness of their answers the SA Test provides a more comprehensive assessment (Figure 1).

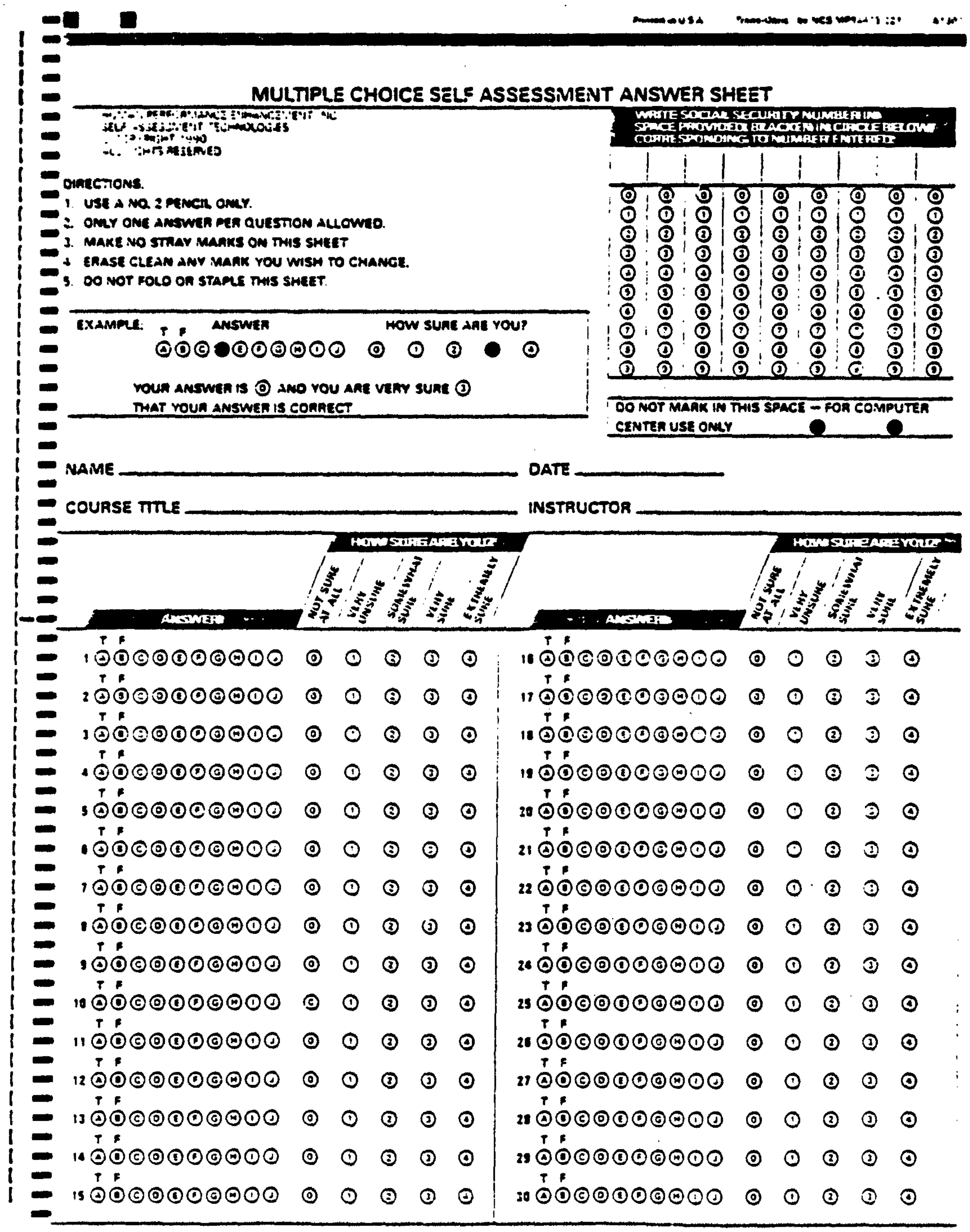


Figure 1. The Multiple Choice Self Assessment Test answer sheet

The test taker indicates on a carefully designed logarithmic five-point scale, how sure he or she is that the selected answer is correct. This helps to remove the limitations of the usual multiple choice test. Self assessment adds a new dimension which measures the person's "usable knowledge" - that knowledge about which the person is sufficiently sure, so that it will be used to make decisions and to take correct and timely actions.

If you will, image this---

> "Help! Help! Warning light No. 6 just came on."
> "Doesn't that mean that the input valve on line 6 is stuck in the open position - and we need to divert the flow just prior to the input junction".
> "Are you sure? I thought warning light 6 meant that we must close the output valve on line 6".
> "Gee! Where is the operating manual".

If this activity took place in the control room of a nuclear power system, or in a high pressure steam plant, the delay in taking correct and timely action could be catastrophic.

People involved in personnel training must often ponder over the effectiveness of their efforts to avoid situations such as depicted here. How can we be sure that the trainee(s) are in fact understanding what is being taught, and will respond correctly under operational conditions. Self Assessment Testing can give you a leg-up on the answer to that question.

Some might say that interpretation of the facial expressions etc. gives an indication of the classes understanding of the material being presented. However, I believe that you will, in most cases, be inaccurate in these observations, and the test results will confirm it.

The unique and important feature of the Self Assessment Test is its ability to detect and identify topics which a person is misinformed. By misinformed we mean that the person is sure of some knowledge or belief which, in fact, is incorrect. The detection of misinformation is of special importance in certification and licensing because the examinee may make decisions and take actions based upon such strongly believed, but erroneous, knowledge.

In the usual multiple choice test if the answer is wrong, there is no way to measure the person strongly believes that it is correct. Thus, there is no way to distinguish between a person who is misinformed and a person who is simply uninformed. However, in the Self Assessment Test, those people who exhibit excessive confidence about an item of knowledge which is wrong, are identified in the computer analysis of their response.

THE SELF ASSESSMENT SCORE

Now that we understand the SA testing method, how do we assess the resulting scores? Since the SA score, like the percent correct score, is based on the 0 to 100 scale, one might treat them in like manner. Remember that the SA scale is logarithmic, where points are gained or lost based on the level of confidence indicated, not on a one for one basis such as in the percent correct score, if a 75 percent correct score is a 'C' or average test score, an SA score of 75 percent indicates that the person is somewhat knowledgeable on the topics of the test, and they are somewhat accurate in assessing their own knowledge. Both scores may be considered a passing grade for a course. However, if the person were being trained to be

knowledgeable in complex systems and be expected to react in a timely manner to contingency situations, it is unlikely that person is qualified. Therefore, acceptable scores should be established based on the complexity of the training required.

The SACAT testing method provides scores for the classes as well as individuals. Therefore, it is capable of providing a measurement of the effectiveness of the course being taught.

It is important to understand that a student or trainee who scores well in the percent correct, but average in SA, that person is assessing themselves too low which indicates that they would fail to take activities and tasks which they, in fact, could perform; which means that they do not reach their full potential. On the other hand, if a person assesses themselves too high, then they tend to take on tasks which produce failures and errors (perhaps with serious consequences); they lack the knowledge they thought they possessed.

As students and trainees learn, by assessing their knowledge, they will also learn to adjust their study habits. They will begin to acknowledge what they know and don't know. In doing so they will gain confidence in the knowledge they possess, resulting in changes in attitude about the learning process.

SELF ASSESSMENT TESTING METHOD ASSISTS THE TRAINER

SA Testing Computer Analysis provides access to a variety of information. Perhaps of the most use to the trainer is the Instructor's summary (Table 1) which includes a list of the specific test items about which the students or trainees are:

a. Misinformed, i.e., those test items (3-10%) about which many of the students or trainees were sure that their answer was correct, but it was wrong. These topics must be addressed specifically to dispel their beliefs.
b. Uninformed, i.e., those items on which many answers of the students or trainees were wrong.

The computer analysis also provides a complete analysis of questions, and if requested by the instructor, results of an individual. The latter is useful in formative evaluation between the instructor and the student or trainee.

All of these analyses are available in the time it takes to scan the answer sheet and have the computer derive the results and print them out (10 to 15 minutes max.). Immediate feedback is important to the student or trainee as well as the instructor.

The benefits of Self Assessment Testing as we perceive them at this time are:

- Obtain a more comprehensive measure of a person's usable knowledge.
- Detect areas of knowledge in which a person is misinformed.
- Identify test items which may be misleadingly constructed.
- Encourage more effective study.

INSTRUCTOR'S SUMMARY OF TEST RESULTS

TEST: Psychol 201 Pretest INST: D.P.Hunt DATE: 1 Aug 1990

OVERALL

A perfect score is a 100% Correct and a 100% Self Assessment (SA) Score. On this test : (a) the 75% CORRECT indicates that the test takers are somewhat knowledgeable on the topics of the test and (b) the 75% SA Score indicates that they are somewhat accurate in assessing their own knowledge.

The 46% SURE-and_CORRECT answers for the group indicates a somewhat low amount of usable knowledge. Of the correct answers, the test takers were sure of 61% of them, which indicates that they are inaccurate in the identification of the correct knowledge which they do possess.

TEST QUESTIONS RECOMMENDED FOR REVIEW

a. MISINFORMED. On 5% of their answers, the test takers were SURE that their answers were correct, but they were WRONG. This indicates that they possess a low amount of misinformation about the topics of the test. Test items which stand out because they have a relatively high percentage of SURE-but-WRONG answers are listed below.

Question	Percent Sure Wrong	Correct Answer	Most frequently chose Number Sure-but-Wrong Answer
1	17	B	D
6	10	B	C
7	13	D	C
23	10	A	B
28	23	A	C

b. UNINFORMED. Questions which were answered correctly by fewer than 50% of the test takers are listed below.

Question Number	Correct Answer	Question Number	Correct Answer	Question Number	Correct Answer	Question Number	Correct Answer
4	D	27	D	40	B	50	C
11	C	29	A	41	C	51	B
13	D	30	D	43	A	53	A
15	D	31	B	44	A	55	A
16	B	32	C	45	D	56	C
22	C	33	B	46	B	57	B
24	D	38	C	47	A	58	C

TEST TAKERS WHO ARE ACCURATE IN SELF ASSESSMENTS

The test takers listed below were especially accurate in the self assessments of their own answers. They could be rewarded by, say, adding 3 percentage points to their % Correct Score.

ID Number	ID Number	ID Number
....58532	99938	26779
....58580	12464	58529
....52515	44574	58558

Table 1. A summary of the test results which identifies for the instructor the test items about which the students are misinformed and uninformed.

- Make people aware of self assessment as an important part of their performance.
- Provide practice with feedback in making self assessments.
- Enhance learning.

- Make testing a learning experience.

- Make testing and learning more satisfying and enjoyable.

WHAT IT TAKES TO USE THE SELF ASSESSMENT TEST METHOD

SACAT or SACAT Scan can be incorporated into your program without redesign of the program or disruption in the methods employed.

SACAT -
- An IBM or compatible computer with word processing capability.
- A printer, if printouts are required.
- SACAT Software Package.
- Users Guide.

SACAT Scan -
- An NCS OPSCAN-5 Optical Dual Head Scanner and Scan Tools Software Package (Being adapted for Scantron Model 8400).
- An IBM or compatible computer with word processing capability.
- A printer.
- SACAT Scan Software Package.
- Users Guide.
- Multiple Choice Self Assessment Answer Sheets.

For copies of this article send request to either of the authors at: HPE, Inc., 345 North Water Street, Las Cruces, NM 88001; Tel: (505) 524-4588 or FAX (505) 646-6218.

How Personnel Testing is Adapting

to a Changing Air Force

Ms Tina S. Ziebell
Air Force Military Personnel Center

Abstract

The Air Force relies heavily on personnel testing as a method of selecting, promoting, and identifying individuals with the necessary aptitudes, skills, and abilities to perform well. This function has become increasingly more important as the force undergoes changes and draws down--it is imperative to make the best possible personnel decisions in support of mission accomplishment.

The purpose of this paper is to examine the changes taking place in Air Force personnel testing to accommodate this new environment. To partially compensate for the diminished numbers of personnel, we are taking every opportunity to increase efficiency across-the-board by revising procedures, consolidating functions, and realigning responsibilities. Integrating new technology is a key element in this process.

Current Organization

The Personnel Testing Section is part of the Air Force Military Personnel Center (AFMPC) at Randolph Air Force Base (AFB), Texas, the operational hub for the AF military personnel system. Currently, the Personnel Testing Section is in the Field Activities Division of the Personnel Operations Directorate. The Section is responsible for oversight of 600 test control officers (TCOs) who administer 700 different tests to more than 275,000 personnel annually. These tests fall into three categories: Personnel Procurement, Aptitude, and Interest Testing; Personnel Promotion Testing; and Personnel Proficiency Testing. The Section is also responsible for monitoring research and development of new testing programs, providing guidance on new technology for testing programs, and maintaining accountability and control of test material throughout the Air Force.

The Personnel Testing System provides appropriate instruments to measure aptitudes, knowledge, and other abilities of AF applicants and personnel. Testing saves money and resources by aiding in selection of the best people to access, train, and place in specific jobs. It also assists in identifying the most qualified airmen for promotion and identifies members who have reached a level of proficiency required for some special area (for example, a career specialty or a foreign language).

Changing Role of Testing in the Air Force Environment

The dynamic world environment, including decreases in the AF budget and force, has necessitated we do business more efficiently. We have accepted the challenge and are stepping out as discussed below.

The Weighted Airman Promotion System (WAPS) uses promotion tests as one component to compare promotion eligibles with their peers and determine who should be promoted. AF members competing for promotion to staff sergeant through chief master sergeant are compared using weighted factors. These include experience, performance, and knowledge measured through specialty knowledge tests and general AF/military knowledge tests. All members are tested once per year and are considered for promotion during annual promotion cycles. The WAPS has been in effect for almost 20 years and is well accepted and understood by the Enlisted Force. Because the system has worked well over the years, changes are not easily approved or accepted. Of course, integrity of the system and fairness to individuals are always of the utmost importance. However, procedural changes to the system have been required to keep pace with events.

During Operations DESERT STORM, FIERY VIGIL, and PROVIDE COMFORT, circumstances dictated a deviation from our normal approach. When fighting in the Persian Gulf began, testing was suspended for those personnel in the area of responsibility (AOR); however, testing for all others remained intact. As personnel returned from these contingencies, they were given, as an exception to policy, up to 75 days to study and take care of personal concerns. Under normal circumstances, personnel who are unable to test before or during their TDY are tested as soon as possible upon return. This policy for extended study time created some strange situations that were handled on a case-by-case basis. After the fighting was over and the peacekeeping force began rotating in and out, the policy was questioned many times. When does the extended study time end? How are these individuals different from others who go to less-than-desirable locations to perform their duties and are given no special considerations?

When Mt Pinatubo erupted in the Philippines, similar circumstances arose. People were evacuated quickly--their lives disrupted. Others were required to stay and help with the evacuation, protect American property, and/or clean up. Again, extra time was given to these personnel and many individual circumstances had to be considered.

We recently instituted these same procedures for those personnel who were reassigned from Homestead AFB, Florida. Because of the disaster Hurricane Andrew created and the subsequent disruption of lives of AF members, we allowed up to 75 days for personal and study time upon their reassignment from Homestead. These procedures ensure that members in undesirable situations, whether in combat or a natural disaster, are given a fair and equitable chance to compete for promotion with their peers who did not experience these unique situations.

As the contingencies settled and the urgency abated, other issues surfaced. Is it fair to continue allowing extra study time to these personnel after the threat of combat has subsided? How is the transition made back to normal testing procedures? Will members be disadvantaged upon return to normal testing policy? As a result, current policy for testing members on extended TDY was re-examined. Major command opinions were solicited on how to best handle these types of situations. Responses indicated a desire to allow commanders more flexibility in accommodating situations where their people need special testing considerations. As a result, the procedures allowing

extra personal and study time upon return from TDY have now been discontinued with the exception of a few locations in which conditions still warrant the extended preparation time. These changes will place more responsibility on commanders--they should best know the situation and how it affects their airmen.

Another change to WAPS is the Staff Sergeant (SSgt/E-5) testing and promotion method. All grades, except SSgt, are tested and promoted once a year. SSgts have traditionally been tested and promoted twice a year because of their large numbers. Offsetting this rationale, it is extremely manpower intensive to conduct testing and run promotions for the same group of individuals twice per year. With the reductions in manning of AF personnel offices, we needed to improve efficiency. The system was reviewed and the decision made to eliminate one SSgt testing cycle per year. Although this will create a larger testing requirement within one cycle, manpower and computer resources will ultimately be saved, and a redundant cycle eliminated. Additionally, the requirement to develop and produce two versions of the SSgt test will be eliminated. This alone will shave time off the development effort of the subject-matter experts and save the costs associated with printing and distribution of additional test booklets. All required changes have been made and publicized for an annual testing/promotion cycle for SSgts effective in 1992. These changes are expected to save the Air Force approximately 30% in manpower and computer resources. As one might expect, the transition period will cause some minor problems, but these are being worked (for example, understanding by the Enlisted Force about how the new method will work; delayed testing and promotion opportunity for some members during the transition).

WAPS has traditionally focused on individual effort and contribution in performance and knowledge. As a result, WAPS testing is predicated on a system of self-initiative which includes individual study and preparation. This prohibits any form of group study. This prohibition has increased test compromise rates and presented difficulties in defending its legalities and justifying why it cannot be used for enlisted promotion testing when it is an accepted and effective technique at all levels of education. As a consequence, the Testing Section established a working group to look at the issue of group study from top to bottom to determine if any changes to the current blanket prohibition against group study are warranted. Some questions to be answered are: 1) Are we handcuffing our Enlisted Force by prohibiting a basic study technique used pervasively at all levels of education? 2) Are we hindering our people's opportunity to improve job knowledge and broaden themselves professionally? 3) Are we hindering our people's job performance? and 4) Are we indirectly hurting AF efficiency? A game plan has been established with a corresponding timeline. We have concluded the research phase--searching literature, discussing, and soliciting views from experts and the other Services. Our next step is to survey the Enlisted Force to determine how they feel about the current policy and possible changes. Implementation issues associated with a policy change will have to be explored. Whatever the final outcome of the working group, any change to the promotion system must preserve its integrity, objectivity, and acceptability by the Enlisted Force.

We are also taking positive action to improve the current WAPS study reference distribution system. The primary study reference for WAPS specialty knowledge tests is career development courses (CDC). These are developed for most AF specialities and are used for training and testing purposes. The current distribution ratio for the CDCs to eligible members for promotion

testing is 1:5. Our plan is to increase this ratio to 1:1 so each member eligible for testing may have his or her own set of CDCs. Although this system will initially cost more than the current method, the subsequent availability to airmen will decrease the problems associated with a reduced ratio and provide all individuals with the opportunity for equal access to CDC study reference material. Implementation is expected in August 1993 for promotion testing beginning in January 1994.

Proficiency tests are critical to mission accomplishment. For this paper, discussion will focus specifically on the Foreign Language Proficiency Pay (FLPP) Program. FLPP is essential to the AF readiness capability. Language testing is conducted to identify resources proficient in a certain language. AF members are provided monetary incentives to encourage maintenance of their language skills. Once identified as proficient through testing, they are put in a pool for possible duty should a contingency develop where their language skills are required. Due to budget constraints in 1991, changes were made to the FLPP program to scale down the pay and language requirements. At that time, FLPP was being offered to any member who could achieve proficiency through testing. As a result, the number of qualified linguists far exceeded any AF readiness requirements. This problem dictated quick changes. While working these changes, procedures were developed to streamline the testing function to be less manpower intensive. Increased automation of pay was also accomplished. Language testing was decreased to one cycle per year (vice two) and major changes to the military pay system were formulated to increase automation. These changes have benefited all individuals involved in the administration of the program by decreasing the workload 25% and FLPP expenditures $3.5M as compared to FY 91. These changes have especially helped AF personnel offices, the key players in the whole process, by not only cutting down the testing required but by more evenly balancing the testing workload throughout the year between promotion testing and language testing. This will boost the accuracy and security of each testing program and benefit the AF member who is paid FLPP through more timely testing and more accurate pay transactions.

More Efficient/New Technology

Training individuals to be officers and/or pilots is a very expensive effort. Selection procedures then become critical to getting the most for our money by predicting successfully those candidates who will complete training and do a good job for the Air Force. To increase our success rate in this arena calls for adding new technology and making our current systems more efficient. To do this, changes are being made at this time and planned for the future.

With fewer pilot authorizations available and less money for training, selection of the very best candidates for pilot training is a must. To increase successfulness of our selection procedures, new technology is being implemented in our testing program early next year. Pilot candidates will be administered a computerized battery of cognitive and psychomotor tests on the Basic Attributes Tester (BAT). The BAT is a computerized test station with control sticks and a keypad so examinees can observe the monitor while simultaneously responding with the control sticks and keypad. The control sticks are used in several batteries to measure psychomotor abilities required for success in pilot training. The Pilot Candidate Selection Method will transmit summary scores over telephone lines to a central processing station at Randolph AFB where they will be weighted with candidates' Air Force Officer

Qualifying Test (AFOQT) composite scores and flying hours to produce a percentile score. This percentile score will reflect the ranking of each examinee among all previous pilot candidates who have taken the BAT. It is estimated that $1M a year will be saved through this process.

The AFOQT, our primary officer selection instrument, has also undergone procedural changes. Previous retest procedures hindered our ability to identify individuals most likely to complete an officer commissioning training program. Research indicated the first or second administration is the most predictive of success in an officer training program. Previous policy required individuals to retest if their last set of scores was more than 2 years old and allowed two retests. This obviously mandated unnecessary retesting in many instances. For example, an individual has taken the AFOQT twice and the last administration is over 2 years ago. Procedures required a retest even though the last administraton would provide good information for making a selection decision. Changes were effective this summer to reflect only one retest after the initial administration and levy no time limit on the life of the scores. This ensures AF selection decisions are made with the most valid data while also saving time and money in reduced test administration and preventing forced retests due to retest procedures.

The scoring system for the AFOQT has recently been revamped. The previous scoring system was based on computer technology from the 1960's and required many manual intervention steps. AFOQT scoring procedures have been updated and automated using in-house capabilities. It now requires approximately half the manpower and computer resources to process scores. Our next step is to completely utilize the capabilities of PCs—accomplish scanning and scoring of answer sheets with a PC vice the main frame computer. This will further save manpower and resources required to perform these procedures.

Looking ahead to the future, we realize computerized testing will become increasingly important. We are moving toward this goal by using the BAT to lead the way. With this system already in place, adding the AFOQT to the computer would be a logical next step. We have already planned for this in our budget and are discussing future implementation with our systems experts. We envision the AFOQT will be administered on a computer, responses transmitted over phone lines, and scored automatically at a central site. The number of machines available will need to be increased to accommodate the number of candidates and the time it takes to administer the test. An analysis will be conducted to determine feasibility and cost effectiveness.

Digital response testing may be an interim step before full implementation of computerized testing. This type testing automates examinees' responses to test items through the use of a hand-held, wireless testing pad instead of marking response options on an answer sheet. They are designed to work with tests requiring a single response, e.g., multiple choice. Each testing location would need a PC to transfer responses from the response pad to a floppy disk. The floppy disks containing item responses could then be mailed to the central scoring facility at AFMPC. It may even be possible to transfer the data directly to the scoring site. The advantages are obvious--elimination of paper answer sheets and cost savings associated with their printing, distribution, subsequent mailing to the central scoring site, and storage. The possibility of test compromise may even be decreased. This is an option for the near future for AF promotion testing; however, many issues must be explored before implementation--security, cost, and testing effects.

Down the line, as our experience base is built in both military and civilian testing communities, we will be assessing whether computerized promotion and proficiency tests are feasible. The "how to's" of each will need to be determined, but even more importantly, the effect of computerized testing on the equity and fairness of the testing system will have to be thoroughly examined. We will be working hand-in-hand with the research experts in our efforts to pursue an operational testing program of this nature.

The AF Personnel Testing Function will continue to strive for improvement and look for more efficient ways of doing business. We definitely reject the notion of "that's the way we've always done it" as a reason for doing something--yet we will not change just to appear we are using the latest technology. We've made significant progress recently, but still have much to accomplish. Our charter will be to ensure that any and all changes made will benefit all AF personnel and decision-makers while maintaining the integrity and fairness of the testing programs.

PROBLEMS OF EXAM COMPROMISE

by
L. J. Paquette
R. E. Doucette

Naval Education and Training Program
Management Support Activity
Pensacola, Florida 32509-5000

ABSTRACT

Exam compromise has not received very much attention in the past but has continued to remain an issue of concern both in the military and civilian sectors. The purpose of this paper is to provide an introduction to exam compromise and its associated problems.

After presenting some definitions, the exam compromise source identification process is discussed and with the use of statistical data and other visual aids, exam compromise examples are presented. Finally, some advantages and limitations of the exam compromise review or screening procedures are discussed.

SOURCE IDENTIFICATION

The primary beginning process in the investigation and discovery of compromise on the advancement examination is the extreme high standard score of 80. Standard scores range from 20 to 80 with 20 being the extreme low standard score. Candidates who attain the extreme high standard score of 80 are initially investigated by Naval Education and Training Program Management Support Activity (NETPMSA). It should be noted that approximately one candidate in every five hundred falls into this category. The investigation involves a review of the candidate's past exam history, length of service, time in paygrade, performance mark, and scores from other candidates from his activity. The candidate's prior examination history is an excellent indicator of his or her present or future success.

The reliability of most advancement exams is quite high, and the probability that a candidate would drastically increase his standard score from one cycle to the next is remote. If a candidate's score increases significantly over his/her previous examination participations for the same exam rate, then the candidate is researched for suspicion of compromise. Many times these increases are valid due to the candidate studying and on the job training within the present rate.

When there is more than one candidate from the same activity who attains a standard score of 80, there is generally a thorough investigation of all candidates participating in that exam cycle from that activity. Candidates with high scores and who also

participated in the same advancement examination from the same activity are matched for significantly high identical wrong response patterns. In essence, both identical right responses and identical wrong responses are investigated.

The sample case study, enclosure (1), demonstrates how this process investigates suspect compromise.

There are secondary processes in the investigation and discovery of compromise in the advancement examination. The model of the advancement examination is the normal bell-shaped distribution. Frequency distributions of all advancement examinations are analyzed for discrepancies such as bimodal distribution. In this case, a cluster of high scores (or low scores) may occur which indicates that this group of candidates is significantly different from their peers. Further analysis of these candidates is conducted using the previously discussed procedures.

Still another source of investigation is the report of a missing exam or other such irregularity in the exam shipping process. Investigations are conducted whenever specific allegations are made that compromise has occured. In some cases the allegations are made by anonymous letter, but all allegations are investigated. There have been some instances where a xerox copy of the examination was made available for sale to potential candidates and eventually confiscated by Naval Investigative Service personnel. Still other cases involved altering of the plastic encasement of the individual examination. Finally, a few cases have occurred where the answer sheet of a consistently high scoring candidate was altered so that he failed the examination. These cases are also investigated as discussed previously.

The primary advantages of reviewing suspected compromise cases are the preservation of advancement exam integrity and fleet morale. The major limitation to the process is the fact that the compromise analysis must commence, in most cases, after candidate answer sheets are received and scored or approximately two months after examination administration. Despite this time lag, the methods discussed have proven to be an effective means of dealing with suspected compromise.

APPLICATION OF A MODEL TO DETECT COMPROMISE ON NAVAL ENLISTED ADVANCEMENT EXAMINATIONS

Dr. Ronald Cody (1985) developed a statistical model at Rutgers Medical School to detect cheating in the classroom. He identified the suspected cheater as student C and the source of the cheater's answers as student S. Based upon the source's set of incorrect test items, he calculated the probability of obtaining the observed number of identical wrong responses for student C (due to chance) by determining the z-score

$$z = \frac{M_c - \bar{M}}{s}$$

where M_c is the number of identical wrong matches for student C with the source, $\bar{M}$ is the expected number of identical wrong matches, and s is the estimated standard deviation of the distribution of identical wrongs. Dr. Cody then computed the probability of attaining at least the number of observed identical wrong matches, M_c, through use of a one-tailed z table.

NETPMSA has been using Dr. Cody's procedure since March 1988 as the primary method in making probabilistic statements regarding the suspected collusion between pairs of candidates participating in Navywide Enlisted Advancement Examinations. A pair or group of candidates suspected of compromise is usually identified on the raw score frequency distribution for an exam rate. Pairwise candidate comparisons with regard to identical responses are then computer generated for each of the zones of suspicion for the flagged exam rate.

For our purpose we set

$$z = \frac{IDW - \sum_{i=1}^{n} P_{yi}}{\sqrt{\sum_{i=1}^{n} [P_{yi}][1 - P_{yi}]}}$$

where n is the number of questions answered wrong by the source, IDW is the observed number of identicial wrongs for candidate C with the source, P_{yi} is the proportion of candidates choosing the same wrong response as candidate C on candidate S's i^{th} wrong answer, $\sum_{i=1}^{n} P_{yi}$ is the sum of P_{yi} or the expected number of identical wrongs, and the quantity $\sum_{i=1}^{n} (P_{yi})(1 - P_{yi})$ is the estimated variance of the identical wrongs distribution between the source and all other candidates for the same exam rate.

For the example given, Candidate A (the source S) made a raw score of 123 on the $XYZ1$ exam while Candidate B (the suspected cheater C) made a raw score of 122 on the same test. The z-score used to determine the probability of Candidates A and B having 20 or more identical wrong responses in common is determined by utilizing the data obtained from the item analysis for exam rate $XYZ1$. The calculated z-score is

$$z = \frac{20 - 7.52}{\sqrt{4.5054}} = 5.87$$

This z-score corresponds to a probability of $P < 2.2 \times 10^{-9}$. Stated another way the probabilty that candidates A and B would have the observed 20 or more identical wrong responses is less than once in every 450 million cases.

In addition, the maximum possible number of identical wrong responses is computed using the following equation:

> 150 (test items) - 118 (identical right responses) - 5 (items that candidate A answered correctly but candidate B missed) - 4 (items candidate B answered correctly but candidate A missed) = 23 maximum possible identical wrong responses

We now have added information that the candidates answered 20 responses identically wrong out of a maximum possible of 23. In many of the cases, in which we are involved, the number of identical wrong responses is the same as the maximum possible number of identical wrongs.

As a result of identifying possible exam compromise, NETPMSA's statistical analysis section prints a report containing a synopsis of the identical response data. In addition, the physical location (command) of the candidates is noted as is each candidate's prior exam records, if any. Prior exam records, along with the information regarding the current exam, give investigation officials more facts with which to make a determination of the case. The report is then sent to Washington who in turn has the command involved conduct an investigation to determine if compromise has occurred. The results of the command's investigation are forwarded to Washington with their conclusions and recommendations. Washington, in turn, notifies NETPMSA as to what action to take. These actions include but are not limited to 1) clearing a candidate from a BUPERS HOLD status, 2) invalidating a candidate's exam results, or 3) sending a candidate a parallel form of the test and monitoring the results; i.e., validating the initial exam results if the candidate scores the same or greater on the parallel test.

REFERENCE

Cody, R.P. (Ed.D.), (1985). COMMUNICATIONS- Statistical Analysis of Examinations to Detect Cheating. Journal of Medical Education. V60 n2 Feb 85, 136-137

SAMPLE **NAVAL EDUCATION AND TRAINING PROGRAM MANAGEMENT SUPPORT ACTIVITY PENSACOLA, FLORIDA 32509-5000** SAMPLE

FOR OFFICIAL USE ONLY

Subj: HIGH EXAM SCORE DISCREPANCY AND RESPONSE PATTERN NOTED ON THE MARCH 1992 ADVANCEMENT EXAMINATION FOR RATE XYZ1

Acty: (UIC XXXXX) NAVAL EDUCATION AND TRAINING PROGRAM MANAGEMENT SUPPORT ACTIVITY PENSACOLA FL 32509-5000

CANDIDATE	SOC SEC NO	EXAM RATE	RS	SS	%TILE
CANDIDATE A	123456789	XYZ1	123	80	99.86
CANDIDATE B	111111111	XYZ1	122	80	99.86

In the March 1992, Series 135 examination for XYZ1, there were approximately 500 candidates. The raw score mean in terms of the number of correct answers was 75. CANDIDATE A and CANDIDATE B were found to have 118 identical right responses and 20 identical wrong responses for a total of 138 identical responses. The probability that these two candidates would answer 20 wrong responses identically (out of a maximum of 23) with their given raw scores is less than once in every 450 million cases. It is noted that both candidates were administered the Series 135 XYZ1 examination at NETPMSA PENSACOLA FL.

PRIOR EXAM RECORDS

CANDIDATE	CYCLE	DATE	EXAM RATE	SS	%TILE
CANDIDATE A	132	9-91	XYZ1	70	97.72
	131	3-91	XYZ1	64	91.92
	128	9-90	XYZ1	60	84.13
	124	9-89	XYZ1	61	86.43
CANDIDATE B	132	9-91	XYZ1	38	11.51
	128	9-90	XYZ1	40	15.87
	124	9-89	XYZ1	33	4.46
	123	3-89	XYZ1	36	8.08

CANDIDATE A has participated in four previous XYZ1 examinations. His scores on these exams are indicative of a high scoring candidate. CANDIDATE B has also participated in four previous XYZ1 examinations. His scores on these exams are indicative of a low scoring candidate. Now in the March 1992 examination both candidates have become outstanding scoring candidates.

It would seem that some factor other than chance has influenced the situation and caused the remarkable performance and response patterns exhibited by these candidates.

FOR OFFICIAL USE ONLY
ENCLOSURE (1)

THE EFFECTS OF UH-1 EXPERIENCE ON UH-60 SIMULATOR PERFORMANCE: A PRELIMINARY STUDY

Charles Salter, John Crowley, John Caldwell, and Ron Smith
United States Army Aeromedical Research Laboratory (USAARL)
Fort Rucker, Alabama

ABSTRACT

The objective of this study was to determine if U.S. Army rotary-wing aviators, not qualified in the UH-60, could be trained to use the USAARL UH-60 flight simulator as test subjects. The initial premise was that basic pilotage skills required for rotary-wing flight were essentially the same regardless of aircraft. Eight volunteer UH-1 qualified aviators were selected as test subjects. Each received 1 hour of ground training regarding UH-60 system operations and a 1.5 hour UH-60 simulator orientation test flight. Then each subject flew eight 1-hour test flights conducted over a 4-day period. The flight profile consisted of 23 standard aircrew training maneuvers (ATMs) required during "check rides" to determine pilot proficiency. Flight data were acquired on a VAX 11/780 computer interfaced to a Perkin-Elmer digital computer which controls the flight simulator. Specific data points collected were magnetic heading, indicated altitude/airspeed, climb rate, turn rate, roll angle, slip ball position, radar altitude, problem/ freeze, ground speed, and bearing/range/time to destination. Preliminary analysis of the results indicates that many UH-1 pilots performed better than UH-60 pilots previously tested in the simulator. On some maneuvers, the subjects showed steady improvement with practice, while on others their initial performance was as good as their final showing, despite having never flown a UH-60 aircraft or simulator before.

INTRODUCTION

Literature searches revealed no studies which directly addressed whether UH-1 pilots could be quickly trained to asymptotic performance levels in the UH-60 flight simulator. Some previous studies (e.g., Ross and Mundt, 1988) placed pilots of one type of aircraft into simulators of a different type, but the effects of other variables rather than the shift itself were assessed. Most related studies among rotary-wing aviators have dealt with transferring learning in the simulator to actual flight, i.e., the validity and usefulness of simulator training. For example, Kaempf and Blackwell (1990) studied 20 aviators in two groups: a control group which trained to proficiency in aircraft while the experimental group trained to proficiency in a simulator and then the aircraft. They found limited, but positive transfer of simulator training to actual flight.

By contrast, Caro (1972) found great positive transfer of training from a simulator to actual flight. The total training time (simulator and aircraft) averaged 49 hours for the test group versus 86 hours under the conventional training program. Similarly, Weitzman, Fineberg, Gade, and Compton (1979) found that

simulators promoted positive transfer in maintaining instrument flight proficiency among experienced pilots.

Farrell and Fineberg (1976) compared one type of flight skill to another--examining whether extensive experience in general flight navigation would transfer to extremely low level flight (nap-of-the-earth or NOE) navigation. The new graduates were not significantly worse than the highly skilled pilots, perhaps because they enjoyed the advantage of recent NOE training. This study suggests that the effects of extensive UH-60 flight experience on UH-60 simulator performance might be achieved by other pilots with only a relatively short period of simulator training.

The previous research which comes closest in design to the current study concerns backward transfer, in which pilots train to proficiency in aircraft and then are tested in simulators. Kaempf, Cross, and Blackwell (1989) conducted a study of backward transfer using the Flight and Weapons Simulator (FWS), finding among the 16 AH-1 instructor pilots a low degree of backward transfer. This suggests that almost any differences between simulator and aircraft may reduce backward transfer, with those most proficient in the aircraft experiencing the greatest initial problem in the simulator.

EXPERIMENTAL DESIGN

Subjects. In this preliminary study subjects were 8 volunteer U.S. Army aviators between the ages of 21 and 40. Four were low-experience pilots with less than 500 hours of flight time and four had greater experience, with 500-1500 hours of flight time.

Flight performance evaluation. All training and testing were conducted at the U.S. Army Aeromedical Research Laboratory (USAARL) facility, using the USAARL UH-60 research flight simulator. This motion-base system includes an operational crew station, computer-generated visual display, environmental conditioning, and a multi-channel data acquisition system. The UH-60 simulator incorporates an automatic flight control system (AFCS) to enhance the static stability and handling qualities. The stabilator is a variable angle of incidence airfoil which enhances the handling qualities. The automatic mode of operation positions the stabilator to the best angle of attack for existing flight conditions. These various systems assist the pilot in holding heading, altitude, rate of turn, airspeed, etc.

Flight data were acquired on a VAX 11/780 computer interfaced to a Perkin-Elmer digital computer which controls the UH-60 flight simulator. This system is capable of monitoring any aspect of simulator control, from heading, airspeed, and altitude, to Doppler readouts, switch positions, and operator console inputs. However, for the purposes of this study, only 17 channels of data were monitored (e.g., heading, airspeed, altitude, climb, slip, roll, turn, aircraft position, and bearing/range/time to destination). The acquired data points were stored on the VAX 11/780 and then

transferred to the main USAARL computer, a VAX 11/785. Flight performance scores including root mean square (RMS) errors were derived from specialized software routines developed at USAARL by Jones and Higdon in 1991.

The flight performance evaluations required the subjects to perform the maneuvers listed in Table 1. The first part consisted of tactical navigation that required the subjects to use visual cues, GPS or doppler information, and time information to correctly navigate the course. The second part consisted of nontactical maneuvers that required the subjects to perform precision maneuvers based on instrument information. These maneuvers are of the type typically flown in a UH-60 aircraft and are described in the Aircrew Training Manual (ATM).

Table 1.
Simulator Flight Maneuvers

	Maneuver	Description
1.	Low Hover	Maintain HDG 150, ALT 10ft
2.	Low Hover Turn	HDG from 150-330, ALT 10ft
3.	High Hover	HDG 330, ALT 40ft
4.	High Hover Turn	HDG from 330-150, ALT 40ft
5.	Navigate to Chk Pt1	Maintain GPS HDG within 10°
6.	Navigate to Chk Pt2	Maintain GPS HDG within 10°
7.	Navigate to Chk Pt3	Maintain GPS HDG within 10°
8.	Navigate to Chk Pt4	Maintain GPS HDG within 10°
9.	Navigate to Chk Pt5	Maintain GPS HDG within 10°
10.	Transition	Establish HDG 360, ASP 120, ALT 2000ft
11.	Straight & Level	Maintain HDG 360, ASP 120, ALT 2000ft
12.	Left Std Rate Turn	360 LSRT (20), ASP 120, ALT 2000ft
13.	Straight & Level	HDG 360, ASP 120, ALT 2000ft
14.	Climb	To 2500ft @ 500 fpm, HDG 360, ASP 120
15.	Rt Std Rate Turn	180 RSRT, ASP 120, ALT 2500ft
16.	Straight & Level	HDG 180, ASP 120, ALT 2500ft
17.	Rt Std Rate Turn	180 RSRT
18.	Climb	2500ft-3500ft @ 500 fpm
		**** Turn AFCS Off****
19.	Descend	3500ft-3000ft
20.	Left Std Rate Turn	180 LSRT
21.	Descend	2500ft-2000ft
22.	Left Std Rate Turn	180 LSRT
23.	Straight & Level	HDG 360, ASP 120, ALT 2000ft
24.	Rt Std Rate Turn	360 RSRT
25.	Descend	2000ft-1000ft

PROCEDURE

Each subject received 1 hour of ground training regarding UH-60 system operations and a 1.5 hour UH-60 simulator orientation test flight. Then the subject flew two 1-hour simulator test flights each day over a 4-day period (8 total flights).

DATA ANALYSIS

The flight performance data were divided into a specific series of maneuvers, and the various control parameters (heading, altitude, etc.) were scored using locally developed computerized routines. The scoring consisted of calculating RMS errors for each parameter from each maneuver, and storing these RMS errors in data files which were subjected to statistical analyses.

In order to calculate RMS errors for each of these parameters, an ideal value was selected against which the actual control accuracy was evaluated. For instance, if a straight-and-level segment was supposed to be flown at a heading of 180 degrees, an altitude of 1000 feet, and an airspeed of 90 knots, RMS errors were calculated by determining the actual control deviations around each of these values for each of the parameters (heading, altitude, and airspeed). Flight data collected from the subjects were analyzed with a series of BMDP statistical programs. RMS errors were transformed into log naturals (a 1.0 was added prior to each transformation to avoid possible problems with zero values) in order to reduce the impact of occasional extremely large error values. Upon completion of data transformations, a series of repeated measures analyses of variance (ANOVAs) using BMDP4V were conducted. When required, simple effects and contrasts were conducted to followup significant main effects and/or interactions.

Data collected from this study were compared to data collected from an earlier study of the Aircrew Uniform Integrated Battlefield (AUIB) conducted during 1990-1991 at USAARL. All seven AUIB subjects selected for this data comparison were qualified UH-60 Black Hawk pilots and flew the exact flight profile used in this study. However, the hover maneuvers were not deemed equivalent because of different simulator scenes available in the two studies, and they are not discussed or analyzed here. Data estimations were completed by using BMDPAM where the means of available data were substituted for missing values. The factors analyzed in the comparison of UH-1 and UH-60 pilots were group, flight, and iteration. The grouping factor was AUIB UH-60 qualified pilots versus UH-1 qualified pilots. The first within-subjects factor (flight) consisted of four levels; flights 1, 3, 4, and 5. For comparability reasons, the first flights from the training days for the AUIB pilots were considered comparable to the last flights from the UH-1 pilots. The iteration factor had a different number of levels depending on how many times that specific maneuver was performed in each profile.

RESULTS

Some of the statistically significant results we found were:

Straight & Level Flight (SL) with Activated AFCS

The three straight-and-level flight maneuvers conducted with

the AFCS (Automatic Flight Control System) engaged were analyzed with a three-way ANOVA (groups x flight x iteration). Results indicated a group main effect on the heading control ($F(1,13)=9.23$, $p=0.0095$) and slip control variables ($F(1,13)=6.21$, $p=0.0270$), with the UH-1 group (means: HDG=0.57, SLP=0.24) performing better than the UH-60 group (means: HDG=0.77, SLP=0.32). Since RMS is a measure of error, lower mean scores represent superior performance.

Straight & Level Flight (SL) with Deactivated AFCS

The one straight-and-level flight maneuver conducted with the AFCS deactivated was analyzed with a two-way (groups x flight) ANOVA. (There was only one iteration.) Results indicated a significant group main effect on the heading control variable ($F(1,13)=4.94$, $p=0.0446$), (AUIB Mean=1.63, UH-1 Mean=1.11), with the UH-1 group again performing better than the UH-60 group.

Right Standard Rate Turn (RSRT) with Activated AFCS

The two right standard rate turn maneuvers conducted with the AFCS engaged were analyzed with a three-way ANOVA (groups x flight x iteration). Results indicated a significant group main effect on the slip control variable ($F(1,13)=4.56$, $p=0.05$), (AUIB=0.43/UH-1=0.36). Further analysis indicated a flight main effect on the airspeed control variable ($F(3,39)=4.39$, $p=0.0094$). Contrasts indicated significant differences between flight 1 airspeed control and flights 3 ($p=0.03$), 4 ($p=0.03$), and 5 ($p<0.01$) airspeed control. Means are listed in Table 2.

Table 2.
RSRT with AFCS Flight Means (ASP)

FLT	Overall Mean	Group (AUIB)	Group (UH-1)
1	1.28	(1.47)	(1.13)
3	1.03	(1.17)	(0.91)
4	1.07	(1.12)	(1.03)
5	1.10	(1.21)	(1.01)

Climb (CL) with Activated AFCS

Two climb maneuvers were analyzed with a three-way ANOVA (groups x flight x iteration). Results indicated a group main effect on the heading control ($F(1,13)=12.57$, $p=0.0036$), slip control ($F(1,13)=12.41$, $p=0.0037$), and rate of climb ($F(1,13)=5.63$, $p=0.0337$). The AUIB group was worse than the UH-1 group on heading (0.75 vs 0.59) and slip control (0.38 vs 0.25), but the opposite was true for rate of climb (4.54 vs 4.76).

DISCUSSION

The UH-1 pilots, who lacked any UH-60 simulator or aircraft experience, performed significantly better on some of these common maneuvers than did the UH-60 pilots. Furthermore, the UH-1 pilots often performed early in their UH-60 simulator practice as well as they did at its end. These results support the initial hypothesis that basic pilotage skills for rotary-wing flight are essentially the same regardless of aircraft. (This is not meant to suggest that UH-1 pilots are qualified to fly UH-60 aircraft without further training, because emergency procedures and tactical operations are different.) An alternative explanation is that the UH-60 pilots were simply a distinctly different subgroup than UH-1 pilots. Perhaps they were so much more experienced with UH-60 flight operations that interference rather than backward transfer occurred in the simulator, making them perform worse than the other group. Or possibly they lacked sufficient motivation due to boredom. The simulator operator's observations during the study suggest that this last is the most likely alternative.

REFERENCES

Caro, P. W. (1972). Transfer of instrument training and the synthetic flight training system (Professional Paper 7-72). Alexandria, VA: Human Resources Research Organization (AD743155).

Farrell, J. P., and Fineberg, M. L. (1976). Specialized training versus experience in helicopter navigation at extremely low altitudes. Human Factors, 18(3), 305-308.

Kaempf, G. L., and Blackwell, N. J. (1990). Transfer-of-training study of emergency touchdown maneuvers in the AH-1 Flight and Weapons Simulator (Research Report #1561). Alexandria, VA: U. S. Army Research Institute for the Behavioral and Social Sciences (AD-A226 360).

Kaempf, G. L., Cross, K. D., and Blackwell, N. J. (1989). Backward transfer and skill acquisition in the AH-1 flight and weapons simulator (Research Report #1537). Alexandria, VA: U. S. Army Research Institute for the Behavioral and Social Sciences (AD A213 432).

Ross, L. E., and Mundt, J. C. (1988). Multiattribute modeling analysis of the effects of a low blood alcohol level on pilot performance. Human Factors, 30(3), 293-304.

Weitzman, D. O., Fineberg, M. L., Gade, P. A., and Compton, G. L. (1979). Proficiency maintenance and assessment in an instrument flight simulator. Human Factors, 21(6), 701-710.

AUDITORY GUIDANCE IN OFFICER LEVEL TRAINING

Raymond O. Waldkoetter and Phillip L. Vandivier

US Army Soldier Support Center
Fort Benjamin Harrison, IN 46216-5530

Students are apt to report feelings of tension or stress that can interrupt learning at all levels of education and training. It is generally accepted that learning is inhibited by stress to the degree of its intensity. Stress as noted will inhibit aspects of learning, and if it is too prolonged poses health concerns as well (McClelland, 1989). If an auditory guidance technique can be applied to lessen tension in the learning situation, then it follows that more effective performance will occur (Waldkoetter & Mulligan, 1978). Particularly, direct changes in test achievement, skill performance, and related attitudes might be experienced if a technology is used to lessen stress and heighten attention. Such a technology exists in the design of special stereo-cassette tapes that provide a relaxed yet attentive state (Monroe, 1982). While some tape data exist suggesting stress is reduced and learning enhanced through scheduled listening (Waldkoetter, 1983), particular positive changes in test achievement, skill performance and related attitudes have not been fully documented, or at least verified in differing academic or training settings.

The Monroe (1982) system as developed relies on audio-stimuli (sound frequencies) to induce a frequency following response (FFR), hearing sound pulses which respond with similar electrical brain signals. Certain sound patterns create states of awareness that will affect perception and behavior (Green, 1973). The sound pulses are further modified through brain wave synchronization of each hemisphere (Oster, 1973) creating another brain signal from sound pulses in each ear. With the sound pulses resonating with like brain signals, states of consciousness occur to enhance behavior of varying kinds. Two prior demonstrations using hemispheric synchronization (Hemi-Sync) in a military setting have indicated acceptance by students and faculty for using such sound tapes without disrupting the academic/training process (Sternberg, 1982; Waldkoetter, 1983).

METHOD

A test unit class was selected to explore tape use with officer-level students for analysis and evaluation. A Public Affairs Officer Course (PAOC) requiring complex behaviors, was selected at the Defense Information School (DINFOS), Fort Benjamin Harrison, IN. The public affairs officers' training and job involves various pressures and skill demands across military community relations, public affairs communication and media, and broadcasting, and could be affected favorably by technology reducing stress and enhancing learning. This test using PAOC #1-91 was considered feasible by DINFOS in view of the uncomplicated technology, no class schedule disruptions, and test objectives. The following three test objectives were to be evaluated:

1. Determine if the auditory guidance (Hemi-Sync) process increases and augments subject-matter learning as reflected by test scores, exercises and related measures.

2. Determine if other learning experiences are positively affected by the use of the Hemi-Sync process as reflected by training exercises and related measures.

3. Determine if other positive behavioral experiences are activated by the Hemi-Sync process as reflected by training exercises and related measures.

Procedures

The training technology was made available to test class, PACC #1-91, in an initial volunteer sample of 23 officer and selected civilian students (13 male/10 female) out of 44 students during 10 October to 14 December 90 for about 10 weeks. It was decided to have students use six tapes to prepare prior to four scheduled course tests and exercises. The six tapes were to be used before study and during study, before and after testing, since the volunteer students were interested in improving overall course performance. All armed forces were represented in the test class with the class divided into two groups, a test and a control group. The control group would receive only faculty counseling as usually given, and the test group would have sound tape exposure and faculty counseling. Another control group was planned for reference using prior graduated classes for course content comparisons. The sound tapes presented some voice instruction and stereo signals to evoke positive responses of attention, concentration, readiness, and relaxation for study and performance. The six tape album was chosen from the Monroe Institute library and is described as producing the supportive responses for progressive accelerated learning (PAL) in an executive context. The principal tapes are Concentration and Retain-recall-release with four others, which provide a blended variety of signals for encoding various responses to affect desired performance in training, sleep, and other activities. The tapes focus attention on given topics, while the student remains aware and relaxed. Selected mixes of sound frequencies, music, and "voice-over" instructions are specifically designed for the tapes.

A tape usage schedule for the test group was proposed to assure a reasonably acceptable level of use for this study. The Concentration tape was to be used during study sessions before actual testing. The Concentration tape probably needed to be used at least three times before each of four test sessions with the Retain-recall-release tape. Students were not "directed" to use the tapes as volunteer participants, but they were encouraged to follow the proposed schedule to profit from potential benefits in performance. When students commented that they were satisfied with the results of any particular tape in achieving improved performance, they were advised they could continue any tape use at their own discretion. Tape use, however, was suggested to continue at a minimal level through the study session for the "10-week" test, concluding the evaluation sequence. Questionnaires were given as applied to the test and control groups. Besides a special Soldier Support Center (SSC) study coordinator and selected DINFOS professional personnel, a Monroe Institute monitor was to be made available to answer study questions and interpret experiences. Students were encouraged to use the tapes according to the procedures given and were instructed as well with tape descriptions. Resources needed to conduct this study were in the form of several supporting DINFOS professional personnel and faculty, an adequate number of stereo players, headphones, and the given Hemi-Sync tapes.

The current PACC class subject-matter tests and related exercises were accepted to validly record the degree of individual and class training achievement. The end-of-course questionnaire had a questionnaire inventory added to account for course and personal reactions to the Hemi-Sync tapes. The inventory portion of the end-of-course evaluation dealt with nine question areas to help determine how the tape use may have effected course performance. Comparisons were made for both internal and external test class and control groups to better estimate the performance effects attributable to the tapes. Percentages of test group responses, the chi-squared (X^2) statistic, and a correlation measure for significant differences were applied in the data analysis.

RESULTS AND DISCUSSION

While class attrition was not a serious problem as in the enlisted broadcasting course (BBC 1-83), 22 of the 23 volunteer officer and select civilian students of PACC 1-91 participated for test and exercise data and were limited finally with only 14 to 16 responses collected per question for questionnaire analysis. Due to several students choosing not to submit complete questionnaire responses and materials, only sufficient records for 14 to 16 test group students could be analyzed for study purposes. Under the circumstances the sample will give an adequate basis for a reasonable number of worthwhile data observations. A strong point in this study may be that the sample group was evaluated under common stressful circumstances rather than trying to draw inferences from a number of widely differing single student examples. Even though the PAL tapes are available through commercial distribution and have proven successful in a self development format, utilizing the tapes in a highly structured training situation appeared to warrant this study effort.

Test Objective 1

As a class PACC 1-91 appeared to have done as well as or even a little better than preceding Public Affairs classes. This class overall received 93.02 (N=44) as the grade-point average (GPA), while the two immediately prior classes received 92.52 and 92.12, respectively. It is not really possible to attribute the slight GPA increase to the effects of the testing process for the Hemi-Sync tapes. Both the test (N=22) and control (N=22) groups did increase with 93.18 (test) and 92.86 (control) over the prior classes. Some key data suggested that the auditory guidance (Hemi-Sync) process did contribute to increasing and augmenting subject-matter learning for the test group. On the four subject-matter tests or examinations on the major course areas the test group did exceed the control in all four cases if only very slightly (93.91/92.36; 92.04/88.86; 87.18/85.36; and 94.27/93.54). The probability of this occurring is significant statistically using chi-squared in that it would only occur by chance less than five times out of 100 for such groups (X^2 (1, N=4) = 4.00, $p < .05$).

Next, GPAs were compared for the principal subject-matter areas of Journalism, Broadcasting, Public Affairs, and Service Unique were the related training exercises are implemented. Only an extremely slight increase could be observed for the test vs. the control group in three of four comparisons

(88.90/88.90; 96.72/96.52; 96.09/95.47; and 93.68/94.96); the observed difference was not significant (X^2 (1, N=4) = 1.00, $p > .05$). It could be instructor observation and subjectivity played a more direct role in this training aspect with greater emphasis on prior "service unique" experience and with civilian test group students being less experienced. The end-of-course questionnaire inventory did offer some related measures as a comprehensive perspective. Test students (N=14 to 16) indicated that they felt the tapes helped them achieve the course objectives. Although a few (4) did not feel the tapes helped, 12 did report they were helped which was statistically significant, and it would be expected that this difference would occur by chance less than five times out of 100 such measures (X^2 (1, N=16) = 4.00, $p < .05$). Even though not statistically significant it is of practical importance to note that a distinct majority believed the tapes helped performance in the instructional areas of Media Relations (66%), Community Relations (66%), and Command Information (67%).

Even where there were almost as many students reporting they were "not helped at all," several still indicated some degree of their performance being helped. Where the GPA for the Service Unique subject-matter area mentioned above was in favor of the control group, 53% of the responding test group did not experience help in performance from the Hemi-Sync tapes. This could suggest that though several were helped (47%) the tapes' effects were not particularly augmenting enough to let them exceed the control group in this specialized performance. One may observe here that such advanced students with higher skills proficiency are less likely to show little if any noticeable change, other than variously augmented experiences where they have heightened awareness of subject matter, psychological processes, and specific task performance. The first test objective, then, had further modest but favorable support as the test group showed 63% believed the tapes improved or did not restrict their GPA. A generalized summary indicated that nearly 67% or six of nine instructional areas were improved and augmented for test students using the tapes.

Test Objective 2

Other learning experiences were positively affected to some degree by the use of the Hemi-Sync process as reflected by training exercises and related measures. The end-of-course questionnaire inventory provided data about this PAOC that involved training exercise experience and related measures. An overview showed a positive evaluation for the course, with test students (75%) reporting the course demands or difficulty required the expected level of effort, and was statistically significant (X^2 (1, N=16) = 4.00, $p < .05$). Depending on one's orientation toward estimating the difficulty of Public Affairs test sessions, 87% of the student responses ranged from "neither difficult nor easy to very difficult." This may show a positive evaluation of the course's subject-matter content and training exercises, since a school will seem more academically challenging and productive if training is not considered "easy." The above difference would be expected to occur by chance less than one time out of 100 such measures for such a group (X^2 (1, N=15) = 8.07, $p < .01$). Journalism assignments were perceived by 93% of the students experiencing the Hemi-Sync tapes as "neither difficult nor easy to very difficult," supporting the challenging training evaluation also, a difference

which would occur less than one time out of 1000 by chance alone (X^2 (1, N=15) = 11.27, $p < .001$).

Nearly 78% or seven of nine course task performances were indicated by majority ratings as improved or positively affected for the Hemi-Sync test students. The tasks of "Memorizing," "Studying," and "Taking tests" were most favorably affected with a majority of 75%, 71% and 67%, respectively, indicating improved task performances. The least positively affected tasks, "Researching" (33%) and "Managing time" (50%) still experienced limited degrees of improved performances. "Writing," "Managing stress," "Speaking," and "Planning/goal setting," respectively, showed positive majority ratings of change of 60%, 60%, 54%, and 53%.

Test Objective 3

Other positive behavioral experiences were activated by the Hemi-Sync process as reflected by training exercises and other related measures. The specific tapes seemed to produce supportive behavioral conditions and states of awareness. Nearly 67% or four of six of the tapes affected positive majority responses with only two giving an indication of "no help at all" for some during the course. Two tapes were analyzed as being decidedly helpful with the test students. The Concentration and sleep induction (Catnapper) tapes proved statistically significant with 80% indicating Concentration and 86% indicating "Catnapper" as providing help through positive behavioral experiences. These tapes offered a degree of help in the PAOC which could be expected to occur by chance less than five times out of a 100 (Concentration; X^2 (1, N=15) = 5.40, $p < .05$) and one time of 100 (Catnapper; X^2 (1, N=14) = 7.14, $p < .01$).

Positive behavioral experiences associated with the tape use were activated in relation to improved test student responses for instructional areas and task performances, which affected training exercise results. Responses to the questionnaire inventory have confirmed some related measures which reflected test student ratings reporting their positive behavioral awareness and performance reactions. Where the test students responded to whether they experienced any unusual mental and/or physical changes during tape use, 25% reported that they did. The significant difference was in favor of not having such experience, in that it would occur by chance less than five times out of a 100 such measures (X^2 (1, N=16) = 4.00, $p < .05$). But in spite of this difference for not having an experience, it is also operationally significant that some students can experience unusual changes that are personally inspiring.

In Summary

Through the study numerous test group student discussions and comments were exchanged suggesting a largely positive behavioral experience with the Hemi-Sync tapes. Those officer and civilian students participating have individually reported that the Hemi-Sync tapes gave them the sensation of being able to do more in less time and to organize assignments more efficiently. No mention was ever made of tapes adding to the course's learning difficulty, but improved study effort and relaxation did seem to result in the test group students. Because of the course effort and assorted time conflicts most of the

students did not utilize the full six tape album. Several tapes were largely rated as "not at all" helpful or assisting study improvement. This may not indicate tapes were ineffective. It may mean the tapes were not used enough to evaluate them accurately or they did not help performance already at a superior level. Generally, attention and readiness to perform assigned tasks were described as more focused to augment task efficiency. Where a few test students reported negative reactions, they were counteracted by revised tape use and alleviating personal psycho-physical symptoms.

At the end of the PAOC test students (N=16) expressed a substantial relationship between their overall positive evaluation of the course and their belief that the Hemi-Sync tapes improved their overall GPA. The Pearson correlation $(r) = .59$, $p < .01$, being statistically significant, would be expected to occur one time out of 100 by chance for such a group. This validated in part the belief that Hemi-Sync tape effectiveness and course values were related, so that if test students were positive toward course achievement they were also tending to experience positive tape results. The sound technology did appear to favorably affect test and skill performance and related attitudes in this limited study.

REFERENCES

Green, E. (1973). Biofeedback for mind-body self-regulation. In D. Shapiro, T. X. Barber, L. V. DiCara, J. Kamiya, N. E. Miller, & J. Stoyva, (Eds.), Biofeedback and Self-Control. Chicago: Aldine Publishing Co.

McClelland, D. C. (1989). Motivational factors in health and disease. American Psychologist, 44, 675-683.

Monroe, R. A. (1982). The HEMI-SYNC process. Monroe Institute Bulletin, #PR 31380H, Nellysford, VA.

Oster, G. (1973). Auditory beats in the brain. Scientific American, 229, 94-102.

Sternberg, D. (1982). Psychophysical HEMI-SYNC measures. Unpublished manuscript. Nellysford, VA.

Waldkoetter, R. O., & Mulligan, J. R. (1978). A Learning-Receptive State as Induced by An Auditory Signal or Frequency Pulse. Paper presented at the 20th Annual Conference of the Military Testing Association, Oklahoma City, OK.

Waldkoetter, R. O. (1983). The Use of Audio Guided Stress Reduction to Enhance Performance. Paper presented at the 25th Annual Conference of the Military Testing Association, Gulf Shores, AL.

Components and Metacomponents of Intelligence Among Navy and Air Force Personnel

Ronna F. Dillon
Southern Illinois University

The importance of adequate prediction of training and job performance remains among the most important goals of personnel psychologists. A well-documented finding is that neither traditional methods nor traditional measures of the ability substrate of training or job performance accounts for satisfactory amounts of variance in criterion measures of interest. The success of personnel selection and classification efforts rests not only on use of psychometrically sound measures of traditional abilities, such as arithmetic reasoning, but rests also on the understanding that a comprehensive model of intelligent training or job performance also must include the information-processing componential and metacomponential substrates of performance.

One way to address this need for better prediction of intelligent performance is to use comprehensive models of intelligence that tap information-processing componential and metacomponential abilities along with other ability dimensions, such as memory. In earlier work, Dillon and her colleagues (see Dillon, 1989, 1991; Reznick & Dillon, 1988) reported that models comprised of information-processing componential, metacomponential, cognitive speed, learning, and cognitive flexibility indices offer a set of aptitudes that, by themselves, account for significant amounts of variance in test batteries of Navy and Air Force aptitude (e.g., AFQT, ASVAB composites), in measures of Navy school performance, and measures of academic achievement. Moreover, the approaches offer significant incremental validity for predicting school performance when used in conjunction with AFQT (Dillon, 1991) and medical school exam performances (Reznick & Dillon, 1988).

Two programs of research designed to address this need for valid paradigms for use in measuring intelligent performance are reported in this paper. Study 1 describes componential and flexibility domains from a larger undertaking conducted with researchers at Armstrong Laboratory. The indices are validated against ASVAB composites. Study 2 describes information-processing componential, metacomponential, flexibility, and learning domains from a larger program, conducted with researchers at the Navy Personnel Research and Development Center. The components, metacomponents, flexibility, and learning indices are validated against AFQT and Navy Basic Electricity and Electronics (NBEES) school performance.

Two information-processing methodologies are used in these research programs, to derive information-processing componential indices. Paradigms involve either the use of psychophysiological equipment to record ongoing information-processing activity during solution of intact tasks (see Dillon, 1989 for a review of work using eye movement measures of information processing) or some computer-administered means of deriving isolated cognitive processes from task items (see Dillon, 1991; Dillon & Harris, 1992) for examples of this approach).

Levels of individual differences in information-processing componential abilities. Both methodologies yield data regarding individual differences in information-processing abilities at five levels: (A) Stages of information processing, including measures of the number of times components are executed, and the latency for each execution of each component; (B) the sequential distribution of processing steps, such as the percentage of components executed in the main stimulus array prior to the first break in ongoing processing to attempt selection; (C) strategies or strategy components, such as image rotation or double-checking; and (D) adaptations over time, including measures of the information-processing substrates of learning and flexibility. Flexibility (i.e., flexible comparison; See Dillon, 1992a) involves the subject's ability to maintain information-processing efficiency as he or she moves from solution of a set of items that are similar in item type and governing inferences, to another item type. Learning indices tap changes in information-processing efficiency over time; (i.e., trials). Data from the stage, sequence, and learning levels of individual differences, and from individual differences in cognitive flexibility are described in this paper. In Study 1, the computerized information-processing testing paradigm is used, while Study 2 involves use of both eye movement and examinee-controlled, computerized information-processing paradigms. In Study 2, eye movement data are collected from two inductive reasoning tests. Similarities in eye movement indices taken from the different tests reflects the robustness of the information-processing phenomena.

Metacomponential operations concern those mental activities in which subjects engage when they are thinking about their mental processes, planning to engage in problem solving, and/or monitoring solution processes. In Study 2, measures of reported use of learning and memory tactics are used to predict AFQT and school performances.

STUDY 1

Method

Sample

The sample was comprised of 467 Airmen, who had completed basic training but had not yet received school assignments.

Instruments

Componential abilities. Isolation of individual differences in the stage, sequence, and flexibility levels of information-processing was accomplished by means of examinee-controlled computer-administered procedure. Subjects processed separate computer screens, each of which contained information about distinct information-processing parameters for a given item. The subject controls the amount of time, number of times and order in which screens are processed Data pertaining to the information-processing substrate of cognitive flexibility also was derived from this technique.

Procedure

Componential abilities. Items were decomposed into information-processing stages and administered using the computerized procedure. Examinees control movement across screens, governing the amount of time, number of times, and sequential distribution of components executed. Flexibility indices also were derived from these items.

Results

Componential Abilities and Flexibility

Information-processing and flexibility componential abilities are combined into models, which are validated against ASVAB. A model comprised on information-processing componential and flexibility indices accounts for 51% of the variance in the general ability ASVAB composite, $F(5, 447) = 29.53$, $p < .001$. The models tested were comprised of indices of encoding/inference, rule application, confirmation, the percentage of total components executed prior to the first attempt at confirmation, and the first flexibility index, tapping encoding/inference components. Cognitive flexibility indices alone were used to predict performances on the Coding Speed and Numerical Operations composite from the ASVAB, $F(5, 447) = 27.07$ $p < .01$.

STUDY 2

Method

Sample

Eye movement data were collected on 68 Navy recruits, all of whom had normal uncorrected or corrected visual acuity. Thirty-nine of these subjects also completed the metacognitive (i.e., tacit knowledge) measure. The multiple-screen computer-administered paradigm was investigated with 33 different Navy recruits.

Instruments

Componential abilities using the eye movement paradigm. Two instruments were used. The first instrument was comprised of 15 figural analogies, in 3x3 format, taken from the Advanced Progressive Matrices (APM; Raven, 1962). The second instrument was the Reasoning Battery (Dillon, 1992). Test items contained verbal and figural analogies and classifications, involving semantic and nonsemantic relations between stimulus elements.

Componential abilities using the multiple-screen paradigm. The verbal subsets from the Reasoning Battery were used.

Metacomponential abilities. A metamemory instrument was used as a measure of metacomponential processing. The instrument required examinees to judge the frequency with which they engaged in various memory-directed tactics for different memory outcomes.

Cognitive flexibilty. Comparisons were made between information-processing efficiency for the last item in each of the four verbal inductive reasoning subsets from the Reasoning Battery compared to the first item in each subsequent set of trial blocks. Eye movement indices were used as data.

Procedure

Eye movement paradigm. Each item was presented on a 35mm slide and administered on a rear-screen projection device. Recording patterns using a corneal reflective eye tracking system accomplished the isolation of: (a) encoding/rule inference, (b) rule application, and (c) confirmation, as well as individual differences in (d) the order in which processing steps were executed, (e) the number of time steps which were executed, and (f) several measures of chunk size. The information-processing substrates of cognitive flexibility also were derived from this recording technique.

Multiple-screen paradigm. The inductive reasoning subsets described for Study 1 were used. Componential abilities were measured using the multiple screen procedure described for Study 1. Cognitive flexibility was measured as in Study 1.

Metacognitive ability. Examinees used a 4-pont scale to rate the frequency with which they engaged in learning and memory tactics.

Results

Componential Abilities

Eye movement paradigm for testing information-processing componential abilities. When validating information-processing indices against the AFQT, data indicate that eye movement measures derived from the Inductive Reasoning subtests account for 21% of the variance in AFQT score, $F(5, 62) = 3.38$, $p < .01$, while measures derived from the APM account for 19% of the variance in AFQT.

Eye scan measures of information processing, collected during solution of the Inductive Reasoning Test, also are validated against three measures of school performance. Moreover, incremental validity studies are conducted to examine variance in Contact Time, Number of Remediations of Instruction Modules, and Final School Average accounted for by information-processing indices above and beyond the contribution made by AFQT. The information-processing componential models contain measures of the number of encoding/inference fixations, number of rule application fixations, percent of encoding/inference fixations, and number of breaks in main array processing to attempt response selection. With respect to Contact Time, AFQT accounts for 15% of the variance in this criterion, $F(1, 55) = 11.30$, $p < .001$. When the first four information-processing indices are added to AFQT, $R^2 = .27$, $p\ (5, 62) = 4.58$, $p < .001$. The increment in R^2 is significant, $F(4, 63) = 2.55$, $p < .05$. Using Remediations needed to achieve mastery of the criterion, AFQT alone accounts for 11% of the variance in this criterion, $F\ (1,66) = 8.05$, $p < .05$, while $R^2 = .37$ when the first four information-processing parameters are added to AFQT, $F(5, 62) = 4.49$, $p < .05$. The increment in R^2 is significant, $F(4, 62) = 3.38$, $p < .05$. Using Final School Grade as the criterion, 2% of the variance in this criterion is accounted for by AFQT alone. Adding information-processing indices, $R^2 = .18$, $F\ (5, 62) = 2.67$, $p < .05$, during solution of the APM. The increment in R^2 is significant, $F(4, 62) = 3.02$, $p < .05$.

Eye movement paradigm used to test information-processing components of cognitive flexibility. In validating flexibility (i.e., flexible comparison) against school performance, the eye movement technique was used to derive the five-variable model, which accounts for 22% of the variance in contact hours needed to pass a module, $F(5, 62) =$

5.08, $p < .05$. With respect to incremental validity, a five-variable model comprised of the first four indices and AFQT score accounts for 38% of the variance in remediations, (5, 62) = 7.44, $p \leq .001$, compared to 15% accounted for by AFQT alone. The increment in R^2 is significant, $F(4, 62) = 6.02$, $p < .05$.

__Eye movement paradigms used to test the information-processing substrate of learning__. An information-processing models of learning accounts for 16% of the variance in AFQT, $F(3, 63) = 3.92$, $p < .05$. Validating models comprised of information-processing indices of learning against measures of school performance, information-processing indices account for 29% of the variance in Contact, $F(5, 61) = 4.98$, $p < .05$, and 43% of the variance in Contact, when AFQT is added to the first four information-processing variables, $F(5, 61) = 9.40$, $p < .001$. The increment in R^2 is significant, $F(4, 61) = 7.49$, $p < .05$. Validating learning indices against Remediations, a five-variable model accounts for 28% of the variance in this criterion, $F(5, 61) = 6.98$, $p < .05$. The increment in R^2 is significant, $F(4, 61) = 6.19$, $p < .05$.
__Computer-administered paradigm for measuring information-processing componential abilities__. Indices of information processing, derived from the examinee-directed computer-administered paradigm, are validated against AFQT score. As an example of mixed (i.e., stage and sequence) models tested, data indicate that 41% of the variance in AFQT is accounted for by a three-variable model, derived from the Inductive Reasoning items, comprised of the number of times encoding/inference information is processed, the number of times inference precedes rule application processing, and the total time spent processing the encoding/rule inference screen, $F(3, 29) = 8.13$ $p < .001$.

__Metacognitive Abilities__

Metacognitive ability first is validated against AFQT. Data indicate that 10% of the variance in AFQT is accounted for by metacognitive ability, $F(1, 37) = 4.25$, $p < .05$. When two indices of metacognitive ability are used to predict number of remediations necessary to pass course modules, the two-variable model accounts for 39% of the variance in remediations, $F(2, 36) = 11.38$, $p < .001$, compared to 34% of the variance accounted for by AFQT alone, $F(1, 37) = 19.13$, $p < .001$. The increment in R^2 is significant, $F(1, 36) = 3.65$, $p < .05$.

GENERAL DISCUSSION

Personnel selection can be accomplished more effectively when greater variance in training or job performance can be accounted for by models of intelligence or aptitude. Toward this end, the work reported in these studies demonstrates that the two methods of extracting information-processing componential data from complex reasoning tasks yield greater predictive validity than traditional psychometric measurement. Under traditional measurement, test scores are used from the test rather than information-processing indices. Moreover, information provided from examination of examinees' strengths and limitations in distinct information-processing activities is prescriptively fertile, making it more appropriate for personnel classification than test score information.

References

Dillon, R. F. (1989). Information Processing and testing. In R.J. Sternberg (Ed). Advances in the psychology of human intelligence, Vol. 5. Hillsdale, NJ: Erlbaum.

Dillon, R. F. (1992a). A componential subtheory of cognitive flexibility. Manuscript in preparation.

Dillon, R. F. (1992b). Psychometric credibility of the Reasoning Battery. Unpublished manuscript.

Dillon, R. F., Harris, C. A. (in press). The validity of information-processing parameters of inductive reasoning. In Proceedings of the Annual Meeting of the Military Testing Association.

Dillon, R. F., & Tirre, W. C. (1988). Componential and metacomponential abilities among Airmen. . Tech. Rep. 88-1. Carbondale, Il: Cognitive processes research program.

Raven, J.C. (1962). Advanced Progressive Matrices, Set II. New York Psychological Corporation.

Reznick, R. K., & Dillon R.F. (1988). Eye movement measures of success in surgical clerkship. Surgery.

Wagner, R. K. & Sternberg, R. J. (1985). Practical intelligence in real-world pursuits: The role of tacit knowledge. Journal of Personality and Social Psychology, 48, 436-458.

MEASURING MARTIAL ATTITUDES: THE MILITARY ETHOS SCALE (MES) IN RETROSPECT

Stephan B. Flemming
Directorate of Social and Economic Analysis
Department of National Defence, Canada

The most influential examination of the attitudes and values of soldiers in Canada was conducted by Charles A. Cotton (1979). Employing the Institution/Occupation model of commitment to duty by military personnel (Moskos, 1986, 1977; Janowitz, 1977; Segal, 1986), Cotton incorporated the Military Ethos Scale (MES) into a survey administered to a sample of Mobile Command[1] personnel. The model proposes that soldiers typically view their service along one of two attitudinal dimensions. Personnel with an institutional orientation respect the classic vocational ethic of military professionalism, involving unlimited liability to duty. Those with an occupational orientation, in contrast, view their service as would employees in the civil labour force, with the demands of service specifically defined by a contract. The MES was developed for use in measuring the concentration of these orientations among military populations. Cotton's analysis of the survey data, in part based upon MES findings, concluded in consonance with civilianization theory that a majority of personnel viewed their service in an occupational manner. In this study, an appraisal of the construction of the MES was conducted using factor analysis with Cotton's original data set. The test result suggests greater support for the traditional ethic of military moral professionalism among Canadian soldiers than was generated by the initial analysis, and proposes an alternate way of interpreting MES data.

The goal of identifying significant factors in the motivation of soldiers has preoccupied many. With the work of Moskos and subsequently others, notably Cotton, analysts have sought to generate empirical models classifying the constituent elements of motivation in military service. The impact of Cotton's work and of civilianization theory in general has been considerable in Canada. The MES was recently used, for example, in an examination of the values of officer cadets at the military college in Saint-Jean (Maillet,1987). The value of this kind of research has been its avoidance of what Kellett calls the "operationally tempting" tendency to "identify a single source of motivation - God, Queen and country, the Party, the regiment, the group, comrades, or whatever" which poorly reflects the complexity of the motivation of soldiers (Kellett,1986:13). The problem of employing empirical techniques on this level, and of survey data in particular, however, is the difficulty of creating measurement tools that may be used with confidence. Many, of course, argue strongly that social and psychological factors cannot be "captured" and meaningfully reduced to empirical scale scores. The debate over the broader validity of such measures aside, techniques exist for assisting in the interpretation of scale measures. One of these techniques, factor analysis, was used in this study in examining the construction of the MES.

---

[1] The army in Canada.

The results of the test are presented in the following manner. Firstly, salient details related to the conduct of Cotton's original data collection are briefly reviewed, as is the process used in preparing the original data set for this analysis. The remainder of the paper is devoted to the test outcome.

Data Collection (1978-1979) and Preparation for Testing

A 68-item survey was administered at selected major bases across Canada to army personnel of all ranks and trades or classifications. The questionnaire initially addressed basic socio-demographic variables such as educational attainment, age, marital status, primary language, and so on as well as variables specific to military life, including rank, years of service, specialized training completed, and type of current posting. The remaining body of the survey was devoted to attitudinal items about military service, asking the respondents to indicate their degree of support for a variety of military practices and traditions, most of which typically demand greater involvement of military members than do civilian occupations of their employees. Most of these were measured with 5-point Likert scales. This portion of the survey also incorporated an existing scale measuring organizational commitment that was modified to reflect the military context.

In preparing the data set for this analysis, it was necessary to alter its composition in several ways. It was not always possible to identify the specific decisions originally made in grouping the data along key dimensions, such as the distinctions drawn among operational and other personnel. Additionally, a variety of small coding inconsistencies were dealt with. Fewer cases were included in the analysis as a result, totalling 1314. The impact of this was tested by replicating several of Cotton's tests; no significant differences were found.

The Military Ethos Scale in Retrospect

A total of fourteen attitudinal items were included in the questionnaire. Respondents were asked to indicate their reaction to a range of service issues; these are listed in Table 1. In constructing the MES, Cotton selected six of these for inclusion. These items were expected to measure two underlying ethos factors. The first of these, the "primacy" of military duty over all other demands, was represented by attitudes toward postings, conflicts between duty and family, and the primacy of operational requirements over the interests of individual members (items 1, 5, and 8 in Table 1). The second, the 24-hour unlimited "scope" of service, was represented by attitudes toward control over off-duty hours, differences in rank after working hours, and the role of superiors in private life (items 2, 6, and 7).

The six selected variables were each coded so that a high score meant high support for traditional vocationalism, with low scores indicating occupationalism. The responses thus ranged from a minimum score of 1 to a maximum score of 5 within each variable. The equation assigning each respondent an MES score summed the six variable scores, creating a scale variable with values ranging from 6 to 30. The MES assumes, as a result, that each of the constituent variables are equally important in contributing to the overall measure; that in adjudicating a soldiers overall devotion to duty, the degree to which service members believe symbols of rank should matter away from work is of the same significance as the belief that

TABLE 1

ATTITUDINAL ITEMS - MILITARY SERVICE VARIABLES

1. "No one should be compelled to take a posting he or she does not want"
2. "What a member of the Forces does in his or her off-duty hours is none of the military's business"
3. "Putting people on charge is a thing of the past"
4. "Military commanders and supervisors should only have operational control over their personnel, with specialists on base having administrative control"
5. "Military personnel should perform their operational duties regardless of the personal and family consequences"
6. "Differences in rank should not be important after working hours"
7. "What a member does in his private life should be no concern of his supervisor or commander"
8. "Personal interests and wishes must take second place to operational requirements for military personnel"
9. "The Forces should encourage military personnel to live on a base rather than in civilian accommodation"
10. "Military service is a way of life and can never be just a job"
11. "I feel very little loyalty to the Forces"
12. "I could just as well be working for a different organization as long as the type of work was similar"
13. "It would take little change in my present circumstances to cause me to leave the Forces"
14. "Often, I find it difficult to agree with Forces' policies on important matters relating to its members"

individual interests should be secondary to the operational demands of units. This assumption leads to a number of further questions when the item distributions are examined in detail. Is it striking, for example, that 70.9% (n=241) of junior other-rank combat soldiers stubbornly insist that their private lives outside the military are their own business, while a full 90.5% (n=48) of senior combat officers insist that it should not be? Or, further, that 73.7% (n=183) of junior other-rank support personnel think that badges of rank should not have lawful authority after working hours, and 74.5% (n=32) of senior support officers believe that they should? In the case of five of the six variables, young soldiers uniformly expressed dissatisfaction for military practices concretely affecting their lives, while their superiors and officers strongly supported them. The data show, in other words, that the people most accountable to several dictates of military traditionalism are least enamoured of them; it is difficult to imagine that it should be otherwise. There was no such split in the responses to the remaining variable, however. Only a minority of personnel in every rank and trade group set their own interests in the broadest sense ahead of the operational requirements of the armed forces.

This was also the case across other variables in the set that were excluded from the MES, several of which measure attitudes toward variables that go to the heart of the Institution/Occupation debate; for example, a majority of personnel in every rank and trade group expressed loyalty to the military, as was the case of those believing that the military can never be just a job. A total of 66.2% (n=225) and 71.2% (n=242) respectively of junior other-rank combat troops responded in a positive manner to these issues. The original finding that attitudinal barriers obtain across rank and trade sectors is true within particular variables, while others indicate a relatively broader consensus along important dimensions. As we will see, these results suggest an important distinction that may be drawn in evaluating the attitudes of soldiers as measured by these data.

Factor Analysis Findings

Factor analysis is a statistical technique used to identify underlying sets of relationships among variables. In many instances, a complex phenomenon cannot be measured adequately with a single item; factor analysis permits the identification of groups of disparate variables which independently measure constituent elements of the same phenomenon. The groups of items expected to measure "primacy" and "scope" discussed earlier are examples of what factor analysis is designed to reveal. In this test, the technique was employed in discovering a way of identifying attitudinal patterns among all the items in the data set, as well as the relative contribution of each variable to the resulting scale.

In the initial output of the test results, there was high communality among all items, with the exception of item 9 regarding military and civilian housing (at .17) which was dropped from the remainder of the test as a result. The factor analysis procedure identified three core factors among the remaining 13 items, explaining 52.4% of the variance. Three factors were extracted, indicating that attitudes along three distinct dimensions were captured by the thirteen attitudinal items. An immanent logic to the factor loading is identifiable and is discussed in the following paragraphs.

Factor 1

The variables loading significantly on the first factor were 1,2,3,4,6, and 7. The items and the factor weights are contained in Table 2. Each of these specific items refers to a concrete aspect of military life. All involve issues with which most if not all respondents will have had actual experience; unwanted postings, the intrusion of military norms and authority into non-working hours, awareness of or participation in a military trial, the endurance of the diffuse powers of superiors in both life and career, and the prolonged loss of control over off-duty time are all universal military experiences. These six items independently measure aspects of the wider issue of the conflict between military role obligations and individual autonomy. More specifically, this factor is a measure of how individual service members feel about the effect military demands have had on their own rights and autonomy along common experiential dimensions.

TABLE 2

VARIABLES LOADING HIGHLY ON FACTOR 1

Item	Factor 1	Factor 2	Factor 3
1. Unwanted postings	.62	.21	.20
2. Control of off-duty hours	.75	.14	.20
3. Military charges anachronistic	.64	-.01	.19
4. Diffuse authority of superiors	.67	.17	.14
6. Rank after work	.78	.12	.14
7. Superiors in private life	.79	.08	.17

Factor 2

The second factor extracted by the test comprised variables 5, 8, and 10. They are shown in Table 3. Each of these items refers to a broad moral prescription for the military profession as a whole. The respondents were asked by these items to identify their support for traditional professional norms of sacrifice and unlimited liability to duty; of service as distinct from mere work, and performance of duty as an absolute necessity regardless of the needs or wants of individuals or their families. This factor measures the attitude of personnel toward the classic principles of military vocational professionalism, which as we have seen demand that individuals accept that the execution of orders, regardless of their cost, is a moral imperative.

TABLE 3

VARIABLES LOADING HIGHLY ON FACTOR 2

Item	Factor 1	Factor 2	Factor 3
5. Duty before family	.00	.76	.11
8. Primacy of the combat role	.18	.75	-.01
10. Military is a way of life	.23	.52	.22

Factor 3

The third core factor comprises items 11 through 14, which are contained in Table 4. All four of these variables measured an aspect of loyalty to service, of normative commitment to the military. The four items were in fact those from the standardised organizational commitment scale specifically incorporated into the survey. Their emergence as a clearly identifiable factor lends substantial confidence to the analysis.

TABLE 4

VARIABLES LOADING HIGHLY ON FACTOR 3

Item	Factor 1	Factor 2	Factor 3
11. Loyalty to the military	.22	.10	.67
12. Forces work same as other jobs	.22	.23	.52
13. Likelihood leave the Forces	.19	.21	.70
14. Reject Forces policies	.11	-.10	.70

Conclusion

Rather than singularly measuring the degree to which service members respect the classic notion of unlimited liability to duty in terms of primacy and scope, the attitudinal items included in the survey measured the respondents feelings about the demands and sacrifices

military service has required of them, their view regarding the importance of classic principles of unlimited liability to duty that have served to distinguish the military from other professions, and the extent of their commitment to the Forces. What the data show, in other words, are both members' attitudes toward the professional values of which the military traditionally demands observance, and also how they feel about the way those very values have affected their lives.

When the distributions of the three scale variables built from the factor weights shown earlier are compared to the MES, notable and consistent findings emerge. Each scale was constructed so that the higher the scale score, in a range between 6 and 30, the more positive the orientation toward that aspect of military life. The mean MES score of 17.6 was significantly lower than all three of the scale variable means derived from the factor analysis; 18.4 ($t=3.63$, $p<.01$), 20.4 ($t=12.67$, $p<.01$), and 19.4 ($t=8.65$, $p<.01$) respectively. In other words, when the factor loadings are employed to construct scale measures, we find that the MES significantly overestimates the extent to which negative attitudes obtain toward military service on the part of the respondents. As well, while a slight majority of personnel scored below the MES mid-point (50.4%), suggesting a preponderance of occupationalism, the proportions falling below the mid-points of all three new scale variables are all significantly lower at the 1% level (45.7%,33.8%, 34.3%).

This paper has reviewed the results of the application of factor analytic techniques to Cotton's influential 1979 data on the attitudes and values of soldiers. The findings suggest that while a range of concrete practices and traditions of service life were viewed unfavourably, particularly by those most accountable to them, support for principles of military professionalism may have been greater among military personnel in Canada ten years ago than was indicated by reliance upon the MES.

References

COTTON, C.A. Military Attitudes and Values of the Army in Canada. Canadian Forces Personnel Applied Research Unit. Report 79-5. 1979.

JANOWITZ, Morris. "From Institutional to Occupational". Armed Forces and Society. Vol.4, No.1, November 1977. Pp. 51-54.

KELLETT, Anthony. Regimental Customs and Tradition. Directorate of Social and Economic Analysis (DSEA/DND) Staff Note 12/86.

MAILLET, L.J. "The Changing Military Values of Canadian Officer Cadets". Proceedings: 29th Annual Conference of the Military Testing Association. 1987. Pp. 49-52.

MOSKOS, Charles C. "Institutional/Occupational Trends in Armed Forces: An Update". Armed Forces and Society. Vol. 12, No. 3, Spring 1986. Pp. 377-382.

MOSKOS, Charles C. "From Institution to Occupation". Armed Forces and Society. Vol.4, No.1, November 1977. Pp. 41-50.

SEGAL, David R. "Measuring the Institutional/Occupational Change Thesis". Armed Forces and Society. Vol.12, No.3, Spring 1986. Pp. 351-376.

STRESS EFFECTS OF ARMY DOWNSIZING ON FAMILY MEMBERS

Michael E. Freville, Ed.D., Counselor
Fort Knox, Kentucky Community Schools

The downsizing of the military population is an issue of vital concern to both the military and civilian sectors, and increasingly so since the end of the Persian Gulf conflict. With the disintegration of the Soviet empire and the failure of communism, the major threat to the United States and the world suddenly became benign. This, coupled with the current state of the federal budget, dictated a need for a smaller overall military strength after decades of maintaining a force capable of handling any perceived worldwide threat, especially that of the U.S.S.R.

As a school counselor in a Department of Defense Section Six school (Macdonald Middle School, Fort Knox, Kentucky) which houses military family students in grades five through eight, the author perceived during the spring of 1992 a change of emotional status in many of the children. Informal chats with a number of them indicated that a common reference was present, such as topics concerning the downsizing of the military and the domino-like effects on the service member and his or her family. In order to make a more scientific assessment of his perceptions, the author developed two questionnaires to study the situation better.

The two separate questionnaires were developed to assess more accurately perceived stress in both children, the service member, and his or her spouse. Both instruments were randomly given to a cross-section of students and their parents. A total of two hundred questionnaires were given to students in grades five through twelve. A total of one hundred and fifty questionnaires were mailed out to service members and spouses. This instrument was mailed to a cross-section of parents with pay grades from E-4 through O-5. The return rate on this population was sixty-six percent with one hundred questionnaires returned. This relatively high return rate indicated a strong interest in the Army's downsizing at Fort Knox.

The attached instrument aimed at parents has the percentages filled in with responses. A total of ninety-five parents indicated they were or anticipated being affected by the downsizing. Of those ninety-five, seventy-eight said they were feeling increased stress on themselves and/or their family members. Sixty-five were feeling very uncertain about the future. Fifty-four spouses of military members said their husbands were feeling distressed or confused about the downsizing; forty one said their military spouse's work performance seemed to be negatively affected. Seventy family

members said they were worried about their future financial condition if the military member was released early. Thirty-nine responded that their children's overall attitude was negatively affected by their perceptions of the downsizing. Thirty- eight saw evidence of their children's grades going down due to effects of the downsizing. Finally, five individuals said they had complete faith in the Army to take care of their family, if they were impacted by the reduction in force.

Parents were additionally given an opportunity to write in any comments they wished. Representative comments follow:

"I wish the Army would treat us as human beings with feelings and communicate with us about just what is going on. I don't like reading about the downsizing in the papers and not hearing anything factual directly from my chain of command."

"The downsizing, while I'm sure a necessity, should be planned out well in advance. Service members should be given all the facts as much ahead of time as possible so they can plan accordingly."

"I hope I will be able to find a decent job if and when I am cut loose from the Army. I am afraid of the civilian market. I have given all I have to the Army and now I don't know what will happen."

"I admit I am on edge at work and home. I know my wife and I argue more lately. The downsizing is on my mind and the uncertainty worries me."

"My children hear my wife and me talk about the downsizing. I know they worry too. This has to affect them. I don't know how to talk to them about this. I don't have any good news for them."

As indicated earlier, two hundred students in grades five through twelve responded to the questionnaire designed especially for them. The majority of respondents were in grades six through nine and were approximately eleven to fourteen years old. The attached questionnaire indicates their percentage responses. The specific indicators are as follows:

One hundred and forty four students said they had been affected in some way by the news of the downsizing. Of those, sixty-two said they experienced sad feelings and ninety-two indicated they spent time wondering what would happen to their parent/s. Fifty-six students pondered what kind of job their parents would get if released from the

Army. Seventy-eight said they wondered where they would move if caught in the downsizing. The effects on school work and peer relations were considerable. Seventy-three said their grades and school work were negatively affected, while sixty-six indicated they were having interpersonal problems with peers. Ninety-five responded that they perceived their parents to be worried about the effects on the overall family from the downsizing, and fifty-three said their parents were arguing about it. Fifty students reported that they were arguing with siblings more lately and thirty-two said they were getting into more trouble at school for talking back or misbehaving. The questionnaire concluded with questions about post schools. One hundred and eighteen students reported they were glad to go to school on post rather than in a civilian school. One hundred and twelve said that their teachers seemed to understand Army kids better than teachers they had had in off-post schools. Finally, one hundred and twenty-two indicated that they were happy to go to school with kids whose parents were in the Army.

The students were also given an opportunity to express any personal sentiments at the end of the questionnaire. Typical comments include:

"My parents talk a lot late at night about what is going to happen to them. I can hear them and they are worried."

"I have gone to both on and off-post schools during my family's four moves. The teachers in on-post schools have always been sensitive to the frequent moves we make."

"My parents are worried about what is going to happen and talk about money problems."

"In my opinion, if the Army would just tell the people who are leaving and when, then a lot of stress would be eliminated."

DISCUSSION:

While this study pertains only to students attending schools at Fort Knox, Kentucky, it is believed that the results can be generalized to a large extent to similar students attending Department of Defense Schools anywhere, both within the United States and all over the world. Children of all ages perceive and experience stress at a different rate than adults. It is obvious from a review of the questionnaire that many students are directly and indirectly experiencing stress due to the topic of downsizing. The author's initial impressions, before the questionnaire was given, seem to be

confirmed. Overall, the majority of reporting students clearly indicate negative effects from the downsizing activities taking place. Even in families where there is no immediate news of a service member being affected by a reduction in force, there is significant stress being reported.

As far as the service members are concerned, they are experiencing an equal, if not more, amount of stress, which is directly impacting their work performance and home and family life. Many are experiencing uncertainty and worry, which left to the imagination, can result in a person "catastrophizing" the actual situation. There is no doubt that the perceived effects of a downsizing are causing tensions between soldier and spouse/children. Additionally, many are concerned about this situation, and want to care for his/her family as the Army, as always, had the expectation that they do just that. And, according to the results, very few soldiers have faith in the Army taking care of him/her self and family if the downsizing catches them.

IMPLICATIONS:

The author believes that the implications of this study are clear to the extent that there really are things that can be done presently to reduce these negative effects already being reported by children and parents alike. Likened to a combat mission or situation, it is known that the more information given in a timely manner to the soldier reduces stress and increases combat effectiveness. Similarly, the more the soldier and his/her family know about if and when the downsizing will affect them, then their stress level is reduced at least to a manageable level. This results in a better atmosphere at home, on the job, and for the children in school. The more facts, not rumors, that are present throughout this downsizing scenario will result in higher morale for the soldier as long as he/she is in uniform. Some soldiers, depending on their MOS or officer skill, will transfer and integrate into the civilian sector more smoothly than others. But many, because of the above factors, will not do so as smoothly. It is especially those who need as much advance notice and help as possible to ease that transition.

The effects on children cannot be overlooked. Many report the positive aspects of going to school in an on-post school where they benefit by having teachers who seem to be especially sensitive to their unique lifestyle and needs. The authorities in positions of responsibility need to take

into consideration the whole family when planning which soldiers will be caught in the reduction in force. As with any father or mother, stresses at work can trickle quite quickly downhill to the home situation. Children are especially sensitive to these stresses since they do not have the capability of many adults in handling them---particularly when they are not getting the full picture.

RECOMMENDATIONS:

It is highly recommended that as many soldiers as possible be told as far in advance as possible they will be released. This will reduce both perceived and real stresses in the soldier and his/her family members which will result in both improved work performance and better family relations. Indirectly, the increased public relations and "good press" the military garners from such a stand can only help its status among the citizenry. Also, children and family members are the forgotten combat multipliers, and the better the morale is of these individuals, then usually the better the morale is of the soldier, resulting in a person who is more combat ready, as he/she knows his or her family is "in good shape".

The positive aspects of on-post schools cannot be over-emphasized. It is the author's experience and observation that teachers who indirectly serve the military by working with family members in on-post schools are espec ally sensitive to the unique needs, lifestyles, and challenges of these young people. This is not an intangible concept but a philosophy and practice that is visible, real, and measurable. In a time of budget constraints that threaten many aspects of our society including the size of the military the relatively small budget of on-post schools and the positive effects they have for both the service member and family member alike cannot be taken lightly. Many times, it is these indirect supportive services to the soldier that make all the differences in his or her overall morale and mission effectiveness. There are two things that can negatively impact a soldier's effectiveness. One, is the uncertainty of his or her job. The other is the happiness of his or her family, and especially in this study, the children in school. On-post schools don't just teach academics, we also support the military service member in a very tangible way which increases mission effectiveness. And, isn't mission effectiveness a core objective of the military?.

PARENT AND FAMILY MEMBER:

AS YOU KNOW, THE ARMY IS PRESENTLY GOING THROUGH A PROCESS CALLED DOWNSIZING. THIS WILL DIRECTLY AFFECT A NUMBER OF ARMY FAMILIES EITHER NOW OR IN THE FUTURE. THE PURPOSE OF THIS QUESTIONNAIRE IS TO ASSESS WHAT EFFECTS THE DOWNSIZING IS HAVING OR WILL HAVE ON FAMILY MEMBERS, ESPECIALLY SPOUSES OF SERVICE MEMBERS. THE QUESTIONNAIRE IS TOTALLY CONFIDENTIAL - YOU NEED NOT GIVE YOUR NAME. A SIMILAR QUESTIONNAIRE IS BEING COMPLETED BY A NUMBER OF SCHOOL-AGE STUDENTS IN THE FORT KNOX COMMUNITY SCHOOLS. THANK YOU FOR YOUR COOPERATION. PLEASE RETURN IN THE ENCLOSED ENVELOPE.

Sincerely,

Michael E. Freville

Michael E. Freville
Counselor
Macdonald Middle School

1. My family is presently or will be affected by the downsizing. 63% Yes 37% No

IF YOU ANSWERED YES, PLEASE GO ON...CHECK AS MANY OF THE FOLLOWING STATEMENTS AS YOU WISH IF THEY SEEM CORRECT FOR YOU.

2. 82% The downsizing has created additional stress on me and/or my family.

3. 68% The downsizing has made me very uncertain about the future.

4. 57% My spouse (service member) seems distressed and/or confused about the downsizing.

5. 43% The downsizing has negatively affected my spouse's work performance.

6. 48% I find that our family argues more lately due to the effects of the downsizing.

7. 74% I worry about what financial condition we will have in the future if my spouse if released early.

8. 41% My children's overall attitude has been negatively influenced by the downsizing process.

9. 40% I believe my children's grades in school have gone down due to effects of the downsizing.

10. 5% I have complete faith in the Army to properly take care of my family if we are affected by the downsizing.

IS THERE ANYTHING ELSE YOU WISH TO ADD? PLEASE WRITE ANY COMMENTS YOU WISH BELOW. THANK YOU FOR YOUR HELP.

STUDENT: AS YOU MAY KNOW, THE ARMY IS PRESENTLY TRYING TO REDUCE THE NUMBER OF PEOPLE IT HAS. THIS IS CALLED "DOWNSIZING". SOME SOLDIERS WILL HAVE TO LEAVE THE ARMY EARLIER THAN THEY WANT BECAUSE OF THE DOWNSIZING. I WANT TO KNOW HOW YOU FEEL ABOUT THIS DOWNSIZING AND HOW IT MAY HAVE AFFECTED YOU. PLEASE ANSWER THE FOLLOWING QUESTIONS. YOU DON'T HAVE TO SIGN YOUR NAME SO PLEASE GIVE YOUR BEST ANSWERS.

1. My parent who is in the Army is definitely affected by the downsizing. 50% Yes 50% No

2. My parent who is in the Army may be affected by the downsizing 58% Yes 42% NO

IF YOU ANSWERED YES TO EITHER OF THE ABOVE, PLEASE GO ON TO THE FOLLOWING QUESTIONS.

3. I have been affected by my parent being caught in the downsizing. 72% Yes 28% No

IF YOU ANSWERED YES, GO ON:

4. I have been affected in the following ways (check as many as you want):

43% Feelings of sadness.

64% Wondering what will happen to my parent.

39% Wondering what kind of job my parent will get in the future.

54% Wondering where we will move when my parent leaves the Army.

51% My school work has been affected and my grades have gone down.

46% I'm having trouble with some of my friends now.

66% My parents seem to be worried more lately about what the downsizing will mean to us.

37% My parents argue because of the possible effects of the downsizing.

35% I argue with my brothers and sisters more lately.

22% I get into trouble at school more for misbehaving or talking back.

82% I'm glad I go to a school on post rather than one off post.

78% My teachers seem to understand Army kids better than teachers would understand us off post.

85% I'm glad I go to school with kids whose parent is also in the Army.

IS THERE ANYTHING ELSE YOU WANT TO SAY? PLEASE WRITE ANY COMMENTS YOU WANT BELOW. THANK YOU FOR YOUR HELP.

THE SOCIO-ECONOMIC BENEFITS OF HOME-BASING OF ARMY UNITS

Hyder Lakhani
U.S. Army Research Institute, Alexandria, VA

Home-basing of Army units is defined as the relocation of a large number of Army units from outside of the continental United States (OCONUS) to continental United States (CONUS). Under home-basing, there will be a 50% reduction in location of the force in OCONUS -- from about 38 to 40% to only 18 to 20% (or from 300,000 to 150,000). Home-basing is also likely to be accompanied by a longer Permanent Change of Station (PCS) location, say, from an average period of three to six years. The economic theory of firm-specific investment in human capital suggests that considerable benefits will be realized from home basing directly by the soldier, and indirectly by the Army.

The economic theory of firm-specific training or investment in human capital states that, in the short term, a firm or an employer may pay a wage that is higher than the value of marginal productivity (VMP) of an employee during the training period (Goldfarb and Hosek, 1976). The firm-specific training benefits the firm which imparts the training and cannot be transferred to another firm. The training imparted to the employee increases the VMP of the employee. A part of this increase in VMP is recovered by the employer in the form of abnormal profits since the employee is more likely to stay with the firm after the training. The higher wage paid by the employer during the training cycle is recovered by the firm from an employee's higher VMP after the training is completed. The longer this recovery period, the greater is the willingness of the employer to train new recruits. Home-basing of Army units with longer PCS is likely to increase firm-specific training given to Army wives because it will increase the training cost recovery period projected by the firm.

Data and Method

Two complementary databases are used for this analysis: the Survey of Army Families, 1987 (SAF 1987) (Griffith et al., 1987), and the 1989 Army Family Research Program, Soldier and Family Survey - Soldier Data File Codebook (AFRP 1989) (Brinkley et al, 1990). Two types of statistical methods are used. The first method consists of cross tabulations of the benefit variables by CONUS versus OCONUS location of spouses (SAF 1987) or the soldiers (AFRP 1989) and a t-test of significant difference between CONUS and OCONUS values of these variables. The second method consists of a system of linear simultaneous regression equations. It is hypothesized that an increase in a soldier's stay at a location will increase an Army wife's earnings. It is also hypothesized that location in CONUS, accompanied by an increase in her earnings, will increase her satisfaction with Army life, which, in turn, will be associated with an increase in

her desire for soldier's reenlistment in the Army.

Results and discussion of soldier benefits

Family income will increase

The proposed increase in PCS stay will increase Army wives' chances of obtaining better paying, career-based, employment instead of the current situation of part-time work and under-employment (e.g., college graduates working for jobs requiring high school education). As noted above, the longer duration of location will induce employers to invest in firm-specific human capital of the wives. The SAF 1987 data were analyzed to estimate a system of equations comprised of four dependent variables. The predictor of spouse earnings included the variable of the number of months spent at the current location, while controlling for several other explanatory variables noted in Table 1. The results in Table 1, column 3, reveal that an increase in stay at the current location by one month (in excess of the average stay of 26 months) increased an enlisted wife's pre-tax earnings by $21 in excess of the mean spouse earnings of $7,948. The results of similar analysis for wives of officers (not shown for brevity) revealed that an increase in stay by one month increased her pre-tax earnings by $39 in excess of an average spouse income of $10,578. An implication of this finding is that the proposed increase in stay at a location under home-basing of units will tend to increase spouse earnings significantly. Extrapolation of average monthly increase in earnings for a 36-month additional period of home-basing resulted in an increase in enlisted wife's earnings by $756, i.e. by about 10%; and that of an officer's spouse's earnings by $1,404, i.e. by about 13%.

Quantity & quality of spouse employment and earnings

The 1989 AFRP survey asked soldiers if their spouses had problem finding employment. The responses ranged from 1=no problem, to 4=severe problem. The results, in Table 2, reveal that Army wives had significantly ($p < .02$) greater problem finding employment in OCONUS compared to CONUS. Therefore, home basing of Army units in CONUS is likely to mitigate this problem. Moreover, jobs obtained by Army wives in OCONUS are likely to be part-time whereas CONUS employment is more likely to be full-time. Table 2 shows that, of all spouses that were working full-time (10-12 months) in 1986, 65% were located in CONUS and only 35% were located in OCONUS. In contrast to this, spouses in OCONUS comprised 54% of all spouses working part-time (1-3 months) whereas CONUS spouses comprised only 46% of this category (Table 2). It must be added that Army wives in CONUS were able to work greater number of weeks for pay and a greater number of hours per week relative to Army wives in OCONUS. Both of these differences were statistically significant ($p <.0004$). Finally, the jobs held by Army wives in OCONUS appear to pay less than the jobs in CONUS, perhaps because of the differential quality of jobs such as career progressive jobs in CONUS relative to OCONUS.

Analysis of the SAF 1987 data reported in Table 2 suggest that, of all enlisted spouses who earned the highest income bracket of "$25,000 and above" in 1986, 70% were located in CONUS and only 30% were located in OCONUS. Similarly, in the second highest earnings bracket of $20,000 to $24,999, CONUS spouses comprised 83% and OCONUS spouses comprised 17%. Therefore, home basing is likely to provide similar higher earnings in CONUS relative to OCONUS.

Satisfaction with Army life will increase in CONUS

The analysis of SAF 1987 data indicated that wives of enlisted soldiers located in CONUS were more satisfied with the Army as a way of life than those in OCONUS. Table 2 shows that, of all spouses who reported that they were "very satisfied", 73% were located in CONUS and only 27% were in OCONUS. The reasons for an increase in an enlisted wife's satisfaction were analyzed in a regression equation model, using the SAF 87 data, with an Army wife's satisfaction with Army life as a dependent variable. The results were as follows: (1) An increase in stay at the current location significantly ($p<.05$) increased her satisfaction with Army life. Every additional stay by a month in excess of the average stay of 26 months was associated with an increase in satisfaction (Likert scale, 1=very dissatisfied, 5=very satisfied) by .005 points in excess of the mean level of 3.54 points. (2) Location in OCONUS was negatively associated with her satisfaction; conversely CONUS location increased the satisfaction level. (3) An increase in a soldier's rank by one level in excess of the sample mean grade level of 6.05 (where E1=1, E2=2, etc.) increased her satisfaction level by 0.15 points above the mean satisfaction level of 3.54. These results are based on statistically controlling for several explanatory variables listed in Table 1. This Table also shows that an increase in an Army wife's satisfaction with Army life significantly ($p<.01$) increased her desire for her husband's retention.

Home ownership in CONUS will increase

The short stay of an average period of three years under the current PCS policy makes home ownership uneconomical, both in CONUS and in OCONUS, because capital appreciation and equity build up over the short period of three years is not sufficient to recover the initial closing costs. Also, many OCONUS countries do not permit real estate ownership to foreigners. Therefore, more soldiers own homes in CONUS relative to OCONUS. Analysis of the 1989 AFRP data, shown in Table 2, indicates that, of all (n = 524) soldiers who owned their homes, 95% were located in CONUS and only 5% were located in OCONUS. Home-basing will increase home ownership on both counts: relocation from OCONUS to CONUS, and longer PCS.

Quality of family life will improve

Table 1

3SLS Regression Results for interdependence of child care used, spouse income, and retention desires of spouses of enlisted soldiers

Predictor/Variable No. Name	Dependent Variable Hours Child care used last month	 Spouse's pre-tax income $ in 1986	 Eq.3 Spouse's Retention
1. Intercept	28.46	-11841.1	-.62***
2. Spouse's pre-tax income in 1986, $	-	-	6.41
3. Child Care used last month, hours	-	36.01***	-.0004
4. Spouse Worked for pay, hours last week	1.39***	-	-
5. Spouse's Months worked for pay in 1986	1.86***	838.02***	.004
6. No. of children less than 6 yrs.	-2.82	-	-
7. Age of youngest child	-6.91	-	-
8. Spouse's satisf. with Army life	-	-	.12***
9. Soldier's years of service	-.21**	-	.003
10. Soldier's rank	7.14***	-	-
11. No. of months at current location	-	21.26*	-.001
12. Spouses perceptions of soldier's career plans	-	-	.71***
13. OCONUS location	-.01*	-	.05
14. Spouse volunteering in Military org.	-.08	-9.1	-
15. Spouse volunteering in Civilian org.	-.34**	16.03	-
16. Spouse age	1.11*	298.85***	-
17. Spouse education	5.95***	-	-
18. Sp. volunteering for home child care	-.68	-	-

Degrees of freedom = 2,116
System R-squared = .46

* = p <.10 *** p <.01
** = p <.05. All significance levels are for 2-tailed test.

Three measures of quality of family life are analyzed. First, it is hypothesized that an increase in the number of nights spent by a soldier away from home during the last six months, whether in CONUS or OCONUS, unaccompanied by spouse, is likely to reflect a deterioration in quality of family life. Table 2 reveals that soldiers in OCONUS spent significantly greater number of nights away from home so that home basing will improve this index of quality of family life. Second, the 1989 AFRP survey asked soldiers about the extent of their worry about family safety when away from home. The values of this variable ranged from 1=least worried, to 5=most worried. Table 2 shows that the soldiers in CONUS worried less (mean=1.89) relative to those in OCONUS (mean=1.96). Third, soldiers in the AFRP 1989 survey were also asked about their satisfaction with quality time spent with children. Table 2 shows that 60% of soldiers in CONUS were satisfied with this index relative to 40% in OCONUS. Thus, home basing will improve quality of family life.

Benefits to the Army

The Army is likely to benefit from home basing on account of five major cost savings: (1) reduction in future costs of PCS moves (from $1.2 billion in Fiscal Year 1992), (2) reduced need for providing child care facilities due to greater availability and use of civilian child care centers in CONUS, (3) decreased costs of providing OCONUS Army schools because of availability of public schools in CONUS, (4) reduced need to racruit and train soldiers because of a potential increase in soldier retention due to an increase in an Army wife's satisfaction with Army life in CONUS relative to OCONUS, as noted above, and (5) reduced cost of providing on-base housing to soldiers who would tend to own and live off-base.

Disadvantages of home-basing

Home basing is likely to have three disadvantages. First, some CONUS posts will be more attractive than others e.g. posts with considerable spouse employment opportunities will be attractive and soldiers would not like to move away from them and conversely for other posts. This problem can, perhaps, be mitigated by developing a market system of exchanges of PCS assignments. Soldiers can be encouraged to advertise their PCS assignment, along with their Military Occupational Specialty (MOS), rank, and desired location(s) in such media as the Army Times. Those who desire to stay at better locations might be willing to pay a price to soldiers who are willing to accept undesirable locations. The Army policy makers should consider accepting the exchanges, subject to its requirements of MOS and other characteristics. Second, Wood (1982) suggests that Army officers' wives who are employed in the civilian sector are likely to be involved more in civilian community and less in the Army network. These spouses will tend to pull the soldiers away from the Army and toward civilian jobs. Hence reenlistment will decrease instead of increasing. The high quality soldiers can,

perhaps, be retained by increasing such economic incentives as the selective reenlistment bonuses and accelerating the speed of promotion. Third, home basing might generate inequalities in home ownership. Officers, who could afford to buy their own homes, might do so and move off the posts. Enlisted soldiers, who cannot afford to buy homes, will tend to concentrate in Army-owned, on post, housing ghettos, unless policy measures are taken to subsidize and stimulate enlisted home ownership.

References

Brinkley, M. , T. Gabel, L. Bunch, and L. Taslet (1990). 1989 AFRP Soldier and Family Survey - Soldier Data File and Code Book. U.S. Army Research Institute Technical Report, August.

Goldfarb, R.S. and J.R. Hosek (1976). Explaining male-female wage differentials for the same job. Journal of Human Resources. 11 (1), Winter, 98-108.

Griffith, J.D., L.S. Stewart, & E.D. Cato (1988). Annual Survey of Army Families: A Report on Army Spouses and Families in 1987. Summary Report. Alexandria, VA: U.S. Army Community and Family Support Center.

Wood, F. (1982). U.S. Air Force junior officers: Changing professional identity and commitment (Unpublished dissertation). Evanston, IL: Northwestern University.

Table 2

Indicators of Socio-economic Benefits of Home-basing to Enlisted Soldiers

(n are in parentheses)

Index	CONUS	OCONUS	p
1. Problem finding spouse employment (1-4)	2.49 (2,015)	2.59 (1,462)	.02
2. Spouse Employment:	65	35	.0004
Full-time (%)	46	54	.03
Part-time (%)			
3. Spouse Earnings ($):			
25,000 & above	70	30	.02
20,000-24,999	83	17	.02
4. Home ownership (%) (n=524)	95	5	.01
5. Spouse "very satisfied" with Army Life(%)	73	27	.05
6. Satisfaction with housing (1-4)	3.37 (4,127)	3.18 (3,098)	.0001
7. No. of nights away from home	29 (2,364)	35 (1,694)	.0001
8. Worry about family safety when away	1.89 (2,697)	1.96 (2,074)	.03
9. Satisf. with quality time spent with children (%)	60 (1,493)	40 (1,493)	.10

SPECIAL FORCES PRIOR SERVICE PROGRAM[1]

Elizabeth J. Brady
U.S. Army Research Institute

Introduction

The Special Forces Prior Service Program, also known as the 18X Program, was established in 1990 by the U.S. Army Recruiting Command (USAREC) as a means of expanding Special Forces recruiting into additional markets. Under this program, soldiers who have separated from any of the armed services and meet certain requirements, including pay grade (E-4 through E-6) and length of separation (no more than 4 years), are eligible to reenlist in Special Forces (SF).

Men who qualify for the Prior Service Program are eligible to attend the 21-day U.S. Army John F. Kennedy Special Warfare Center and School (USAJFKSWCS) Special Forces Assessment and Selection (SFAS) Program at Fort Bragg, NC. In brief, SFAS assesses individuals for physical fitness, effort, ability to cope with stress, leadership qualities, and ability to work on teams. Soldiers who demonstrate a potential for SF are selected to attend the SF Qualification Course.

A major advantage of the Prior Service Program is that it provides a valuable source of experienced soldiers for the SFAS Program. Prior Service soldiers also bring with them diverse backgrounds that make them a unique and potentially important asset. Partly because of their diversity, Prior Service candidates typically prepare for SFAS by attending a two-week pretraining program that covers the basic skills (e.g., ruckmarching techniques) and attitude development that are needed for successful SFAS participation.

The purpose of this report is to assess whether and to what extent Prior Service soldiers differ from Active Duty and National Guard/Reserve candidates on SFAS outcomes.

---

[1]Presented at the meeting of the Military Testing Association, October, 1992. All statements expressed in this paper are those of the author and does not necessarily reflect the official opinions or policies of the U.S. Army Research Institute or the Department of the Army.

Research Approach

Procedure and Samples

The data used in these analyses came from the ARI-USAJFKSWCS FY91 and FY92 SFAS databases which contains military and personal demographic information plus SFAS performance and outcome data on all candidates.

There were two samples drawn from SFAS classes conducted in FY91 and FY92. Both samples were limited to enlisted soldiers. The first sample consisted of data from the first eight classes in which Prior Service soldiers were allowed to participate in FY91, and the second consisted of seven classes conducted in FY92. Table 1 shows the number and percent of Active Duty, National Guard/Reserve, and Prior Service candidates in FY91 and FY92 SFAS classes.

Table 1

Number and Percent of Active Duty, National Guard/Reserve, and Prior Service Candidates in FY91 and FY92 SFAS Classes

COMPONENT	FY91	%	FY92	%
Active Duty	1625	69%	1696	71%
National Guard/Reserve	410	17%	233	10%
Prior Service	329	14%	450	19%
TOTAL	2364		2379	

Results and Discussion

SFAS Outcomes. Candidates have two opportunities to pass pre-requisites; day 1 and day 3 of SFAS. In order to pass SFAS pre-requisites, candidates must score a minimum of 206 points on the Army Physical Fitness Test (APFT), with no less than 60 points on any event (2-mile run, pushups, and situps) scored for age group 17 through 21, and be able to swim 50 meters unassisted while wearing battle dress uniform and boots. Therefore, a pre-requisite drop is defined as a failure on the APFT and/or the swim test. As shown in Figure 1, Prior Service candidates had the highest overall pre-requisite failures in FY91 and FY92 as compared to Active Duty and National Guard/Reserve.

The high pre-requisite drop rate for Prior Service candidates was largely due to swim test failures in FY91. In fact, it was at least double that of Active Duty and of National

Guard/Reserve. This result may be partly explained by the fact that Prior Service soldiers are not required to take a swim test before SFAS.

In FY92, the high pre-requisite drop rate of Prior Service candidates was largely due to APFT failures. Their failure rate was double that of Active Duty. This result may be explained in part by the simple fact that Prior Service candidates are not, or do not have the time to adequately prepare physically for SFAS.

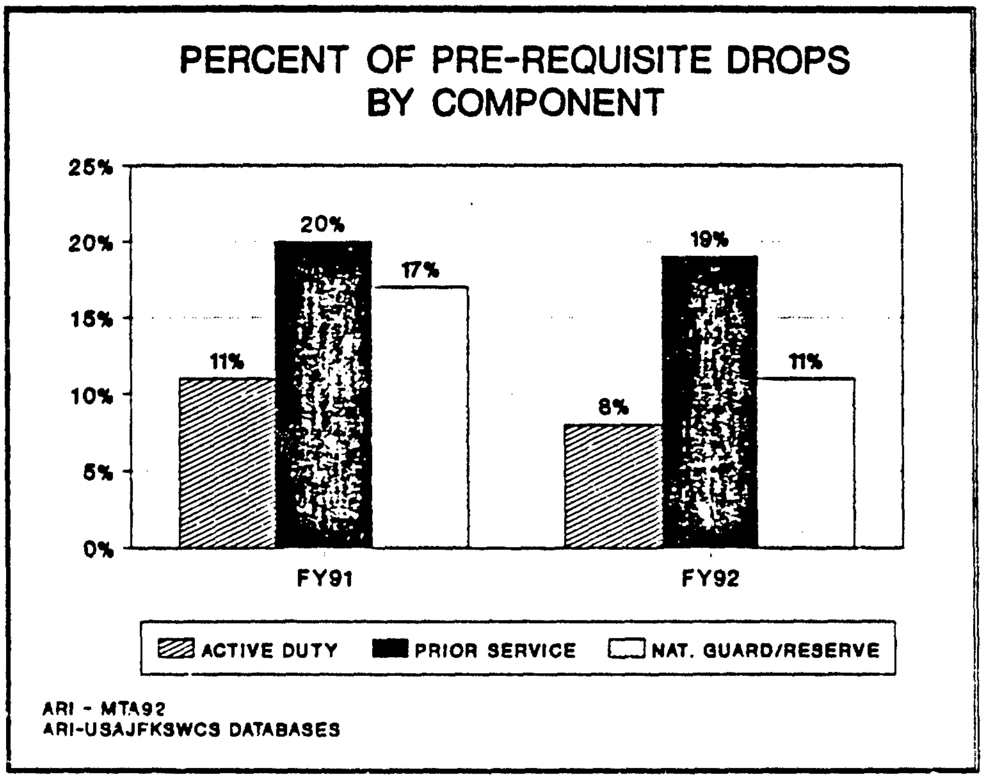


Figure 1. Percent of pre-requisite drops in each component (FY91 and FY92 SFAS classes).

A candidate may not voluntarily withdraw from SFAS until after day 4. Figure 2 shows the percent of candidates who voluntarily withdrew from SFAS. As shown, the percent of voluntary withdrawals for all of the components dropped in FY92. Prior Service candidates had a higher percent of voluntary withdrawals than National Guard/Reserve but lower than Active Duty.

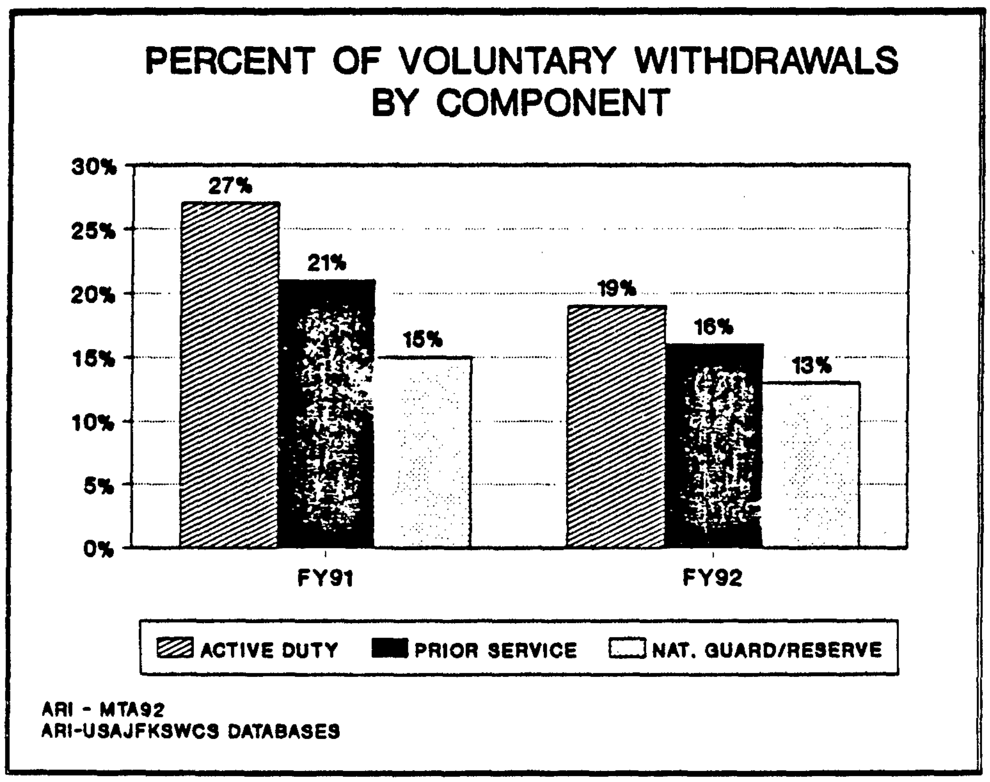


Figure 2. Percent of voluntary withdrawals in each component (FY91 and FY92 SFAS classes).

The third, and most important, outcome examined was the percent of candidates who graduated from SFAS. As shown in Figure 3, the graduation rates increased for all of the components in FY92. Graduation rates for Prior Service candidates was lower than that for Active Duty but higher than rates for National Guard/Reserve in FY91. In FY92, the Prior Service graduation rate remained lower than both Active Duty and National Guard/Reserve rates.

However, since the graduation rate was based on the total number of candidates (including pre-requisite drops), the Prior Service graduation rate is somewhat more impressive than it seems at first glance. Even though the Prior Service pre-requisite drop rate was nine percentage points higher than the rate for Active Duty in FY91 and 11% higher in FY92, their graduation rate was only two percentage points lower in FY91 and eight percentage points lower in FY92. Thus, Prior Service soldiers who passed the pre-requisite tests performed relatively well in SFAS.

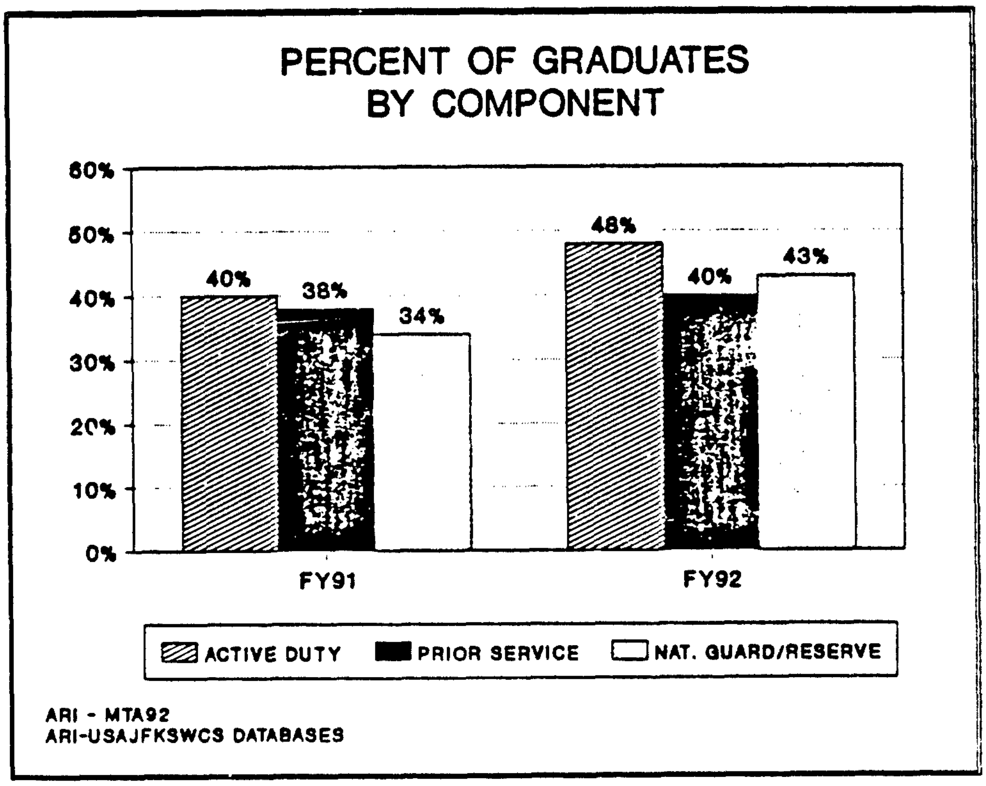


Figure 3. Percent of graduates in each component (FY91 and FY92 SFAS classes).

Recently, USAJFKSWCS and USAREC began to reconsider the Prior Service Program in view of current recruitment needs. After a period of expansion, SF authorizations are now approaching sustainment levels. As a result, the recruiting mission is being reduced (Herd & Teplitzky, 1991), allowing USAJFKSWCS and USAREC to be more selective about personnel applying for SF. For a more indepth look at the Prior Service Program, please refer to Brady and Brooks (1992).

References

Brady, E. J., & Brooks, J. E. (1992). Prior Service soldiers in the Special Forces Assessment and Selection Program: Recruitment issues (ARI Research Report in preparation). Alexandria, VA: U.S. Army Research Institute.

Herd, A. M., & Teplitzky, M. L. (1992). Special Forces recruiting: An overview of current procedures and issues (ARI Research Report 1626). Alexandria, VA: U.S. Army Research Institute.

Demographics and Aptitude

Malcolm James Ree James A. Earles

Armstrong Laboratory, Human Resources Directorate,
Manpower and Personnel Division

The post-war baby boom of the late 1940s, 1950 and 1960s created large cohorts of enlistment and commissioning age. The all-volunteer force of the 1970s was fueled by this group but that is already changing. The future holds few certainties but among these is that the cohort of 18-23 year old eligibles for the service has begun to decrease and will continue to for the next few years. By the middle of the 90s, this cohort will begin to slowly expand. This will have profound consequences which will spread across the entire array of Manpower, Personnel and Training. Recruiting, selection, job classification,and retention will be impacted as the population changes and policies are adjusted to the available manpower supply. Estimates of aptitude distributions, which can be expected to change as the demographics of the population change, are provided through analyses of nationally representative sample of American youth.

METHOD

Subjects

A nationally representative sample of American youth, collected in 1980 by the National Opinion Research Center (NORC), was the sample which formed the basis for these analyses. It serves as the normative reference for the scores for the Armed Services Vocational Aptitude Battery (ASVAB). The sample has 9,173 males and females and in weighted form, has over 25,000,000 subjects (Maier & Sims, 1986; Ree & Wegner, 1990).

Measures

The Armed Services Vocational Aptitude Battery is a multiple aptitude test battery (DOD, 1984) composed of ten subtests. The ASVAB is used by all the Services for enlistment qualification and initial job assignment. The battery has been used in this current configuration of subtests and items since 1980, and is highly reliable (Palmer, Hartke, Ree, Welsh, & Valentine, 1988) and valid (Wilbourn, Valentine, & Ree, 1984). Five measures calculated from the ASVAB are of special interest to the Air Force. The first, the Armed Forces Qualification Test (AFQT) is used by congress to determine relative trainability and quality of recruits for all the services. The other four aptitude indexes are used by the Air Force for classification to jobs. The ASVAB subtests, and composites were described by Earles and Ree (1992).

Procedures

The weights for the sample (which make it nationally representative) correct for over-sampling of ethnic minorities and poor white subjects. Adjustments to these weights were derived from current census estimates and projections (Spencer,1986; Spencer,1989) for the years 1990 through 2010. Two simplifying assumptions were necessary. The first was that the Mid-Series estimate provided by the census was most appropriate. In census projections there are three

estimates of population on the basis of High, Middle and Low Levels of fertility. The Mid-Series estimate was used to avoid extremes. Secondly, it was assumed that within each of the three racial/ethnic groupings ability distributions in the future would remain as in 1980. Another way of saying this is that offspring within a group are expected to resemble the group. No assumptions of genetic inheritance nor environmental influence on test scores are necessary.

There are difficulties in the estimation of statistics for Hispanic population groups. The NORC estimates appear to be based on a definition of Hispanic that was different from that used by the Bureau of the Census. The NORC estimate was that there were 1,544,000 Hispanic youth aged 18 through 23 in 1980. Assuming that deaths approximately equaled net migration for this age cohort, there would be 1,544,000 Hispanic youth 20 through 25 years of age (adhering to the NORC definition) in 1982 (the estimate closest to 1980 in the most recent Hispanic census report). The Hispanic census provides a 1982 estimate of 2,026,000 for "Spanish-Origin" or Hispanic residents of this age group. A multiplier of .767 adjusts 2,026,000 to the NORC value of 1,544,000 and was applied to the census "Spanish-Origin" estimates for 1990, 1995, 2000, 2005, and 2010.

Additionally, net increases from emigration/immigration and the effects of the out-marriage rate (the rate at which members of one group marry or produce children with members of another group) could not be appropriately estimated. It is believed these effects would be small because non-citizens are a negligible portion of the Air Force and out marriage rates tend to cancel across groups.

Distributions of AFQT categories were made. The number and proportion in each category was computed for each year. Additionally, the mean, standard deviation, and selected percentiles were computed for the Air Force aptitude indexes: Mechanical, Administrative, General, and Electronics.

RESULTS AND DISCUSSION

The proportions of racial/ethnic groups in 1980 through 2010 are shown in Table 1. The White proportion of the population falls from 80% to 74% during the period.

Table 1 Racial/Ethnic Percentages of 18-23 Year Old Population

Year	White	Black	Hispanic
1980	80.3	13.7	6.1
1990	78.3	14.6	7.0
1995	76.5	15.4	8.1
2000	76.0	15.7	8.4
2005	74.9	15.7	9.4
2010	73.8	16.0	10.2
1980-2010	-4.5	+1.4	+3.2

Year	Total	White Numbers	Black (x1000)	Hispanic
1980	25,409	20,395	3,470	1,544
1990	22,309	17,478	3,266	1,565
1995	20,399	15,608	3,141	1,650
2000	21,900	16,642	3,429	1,829

2005	23,132	17,336	3,630	2,166
2010	23,265	17,172	3,720	2,373

The Black and Hispanic proportions in this age range increases from approximately 14% to 16% and 7% to 10% respectively. Table 1 also shows the estimated numbers in these groups. The most apparent effect is the consequence of the reduction of number of births--the birth dearth as it has been called in the popular press. The number of military aged young men and women falls by about 3.1 million from 1980 to 1990. In 1995 the number falls further to show a decrement of about 5 million compared to 1980. From 2000 to 2010 the numbers increase but remain 2.1 million fewer than in 1980.

Table 2 shows the distribution of AFQT category by year. Category II, from which the Air Force heavily recruits, shows the largest proportional change.

Table 2 Distributions of AFQT by Category

AFQT Category	%	1980	1990	1995	2000	2005	2010
I	93-99	7.9	7.8	7.6	7.6	7.5	7.4
II	65-92	28.2	27.7	27.3	27.2	27.0	26.7
III	31-64	33.6	33.4	33.2	33.1	33.0	32.9
IV	10-30	21.0	21.4	21.8	21.9	22.1	22.3
V	1- 9	9.3	9.7	10.1	10.2	10.4	10.7
I-IIIa	50-99	51.4	50.6	49.9	49.7	49.3	48.9
IIIb-V	1-49	48.6	49.4	50.1	50.3	50.7	51.1
Estimated Number of Manpower Resources (x1000)							
I		2,007	1,740	1,550	1,664	1,734	1,721
II		7,165	6,179	5,568	5,956	6,245	6,211
III		8,537	7,451	6,772	7,248	7,633	7,654
IV		5,335	4,774	4,446	4,796	5,112	5,188
V		2,363	2,163	2,060	2,233	2,405	2,489
I-IIIa		13,060	11,288	10,179	10,884	11,404	11,376
IIIb-V		12,348	11,020	10,219	11,015	11,727	11,888

Note. % is used to indicate percentile.

From 1980 to 1995 there is a loss of over one and one half million. Category I shows a loss of 450,000 in this same time period. Categories IV and V show increased proportions but a loss of 1,200,000 individuals from 1980 to 1995. However, these two lowest categories climb steadily to 1980 levels by the end of 2010. The Air Force does no significant recruiting in Category IV and is prohibited from recruiting in Category V. The proportion in the upper-half of the distribution (I- IIIa) decreases from 51.4% to 48.9% across the time span estimated.

Finally, Table 3 gives the expected proportions of the population in the four quartiles for 18 to 23 year-old youth for the four Air Force aptitude index composites. Changes in the demographics lead to projections of decreased performance on these classification measures for enlistment age youth through the year 2010.

Table 3 Mean and Standard Deviation and Expected Proportions in Quartiles of USAF Composites

Mechanical	1980	1990	1995	2000	2005	2010
Mean	50.3	49.7	49.2	49.0	48.8	48.4
Std Deviation	28.8	28.9	29.0	29.0	29.1	29.1
First Quartile	26.1	24.6	24.1	24.0	23.8	23.6
Second Quartile	24.3	24.4	24.2	24.1	23.9	23.7
Third Quartile	24.8	25.0	24.9	24.8	24.8	24.8
Fourth Quartile	24.8	26.0	26.8	27.1	27.5	27.9
Administrative	1980	1990	1995	2000	2005	2010
Mean	50.6	50.1	49.6	49.5	49.2	48.9
Std Deviation	29.0	29.1	29.1	29.2	29.2	29.2
First Quartile	26.1	25.7	25.3	25.2	24.9	24.7
Second Quartile	24.3	24.0	23.7	23.6	23.6	23.4
Third Quartile	24.8	24.7	24.8	24.8	24.7	24.8
Fourth Quartile	24.8	25.6	26.2	26.4	26.8	27.1
General	1980	1990	1995	2000	2005	2010
Mean	51.0	50.4	49.9	49.7	49.4	49.1
Std Deviation	29.1	29.3	29.3	29.4	29.4	29.4
First Quartile	26.6	26.1	25.7	25.5	25.3	25.0
Second Quartile	24.0	23.7	23.4	23.4	23.2	23.1
Third Quartile	24.8	24.7	24.7	24.7	24.7	24.7
Fourth Quartile	24.6	25.5	26.2	26.4	26.8	27.2
Electronic	1980	1990	1995	2000	2005	2010
Mean	50.3	49.8	49.3	49.1	48.8	48.5
Std Deviation	28.8	28.9	29.0	29.0	29.1	29.1
First Quartile	25.0	24.6	24.2	24.1	23.8	23.6
Second Quartile	24.5	24.1	23.7	23.6	23.6	23.3
Third Quartile	25.3	25.3	25.3	25.3	25.2	25.2
Fourth Quartile	25.2	26.0	26.8	27.0	27.4	27.9

Note. "Std" is the symbol for standard deviation.

By the year 2010 the Air Force could be much different from now. The proportion of whites will decrease and the proportion of minority group members will have increased. A little over a quarter (26.2%) of the pool of eligibles will be Black or Hispanic as opposed to about 19 percent in 1980. The absolute number of young men and women in the prime recruiting ages will have fallen from the large 1980 value to its lowest in 1995. It then begins a 15 year climb but will still be more than two million below the level of 1980. In the AFQT I-IIIa categories, there will be over 1.6 million fewer young people than in 1980. The Air Force and the other services are currently concentrating their recruiting in these categories and competition can be expected to become more intense. Further, because almost all Air Force officers come from AFQT categories I and II where there will be about 1.2 million fewer in 1995 than in 1980, the recruitment of officers can be expected to become more competitive and difficult.

The four Air Force composites show a decline in average percentile score. In each case the proportion of scores in the two upper quartiles decline while the two lower quartiles increase. The implication is that it will be more difficult to recruit and train individuals for the more difficult technical specialties. For example, the proportion in the first quartile of the Electronics composite drops to 23.6%. This is a drop of over 861,000 young men and women who qualify for training in Air Force jobs requiring first quartile electronics scores. Additionally, industry can be expected

to bid for the services of the highly qualified. The demographic trends suggest that recruitment for the Air Force will become somewhat more difficult. Because aptitude is closely related to training, job performance, retention, promotion and a host of other areas which concern the Air Force the policies of today must be modified to accommodate demographic changes.

REFERENCES

Department of Defense (1984). Armed Services Vocational Aptitude Battery (ASVAB):Test manual North Chicago, IL: U.S. Military Entrance Processing Command.

Maier, M.H., & Sims, W.H. (1986). The ASVAB Score Scales: 1980 and World War II. (CNR 116). Alexandria, VA: Center for Naval Analyses.

Palmer, P., Hartke, D.D., Ree, M.J., Welsh, J.R., & Valentine, L.D. (1988). Armed Services Vocational Aptitude Battery (ASVAB): Alternate Forms Reliability (Forms 8, 9, 10, and 11) (AFHRL-TP-87-48). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Ree, M. J., & Wegner,T.G. (1990). Correcting differences in answer sheets for the 1980 Armed Services Vocational Aptitude Battery reference population. *Military Psychology, 2*, 157-169.

Spencer, G. (1986). Projections of the Hispanic population: 1983 to 2080. Current Population Reports, Series P-25, No. 995, U.S. Department of Commerce, Bureau of the Census. Washington D.C., U.S. Government Printing Office.

Spencer, G. (1989). Projections of the population of the United States, by age, sex, and race: 1988 to 2080. Current Population Reports, Series P-25, No. 1018, U.S. Department of Commerce, Bureau of the Census. Washington D.C., U.S. Government Printing Office.

Wilbourn, J.M., Valentine, L.D., & Ree, M.J. (1984). Relationships of the Armed Services Vocational Aptitude Battery (ASVAB) forms 8, 9,and 10 to Air Force technical school final grades (AFHRL-TR-84-8). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

ASSESSING THE COACHABILITY OF PROJECT A SPATIAL TESTS: DEVELOPING STRATEGIES[1]

Dale R. Palmer
U.S. Total Army Personnel Command

Henry H. Busciglio
U.S. Army Research Institute

Introduction

A number of Project A spatial tests - Assembling Objects, Figural Reasoning, and Orientation - have been included in the Enhanced Computer Administered Test (ECAT) program and are being considered for addition to the Armed Services Vocational Aptitude Battery (ASVAB). Therefore, it is important that these tests remain valid incremental predictors (over and above ASVAB) of various job performance criteria. One important potential threat to the long-term validity of these tests is the possibility of confounding true spatial ability with differential practice and coaching effects.

The following paper describes stages in the development of strategies and scripts for a coaching experiment involving the three spatial tests. Specifically, the stages involved in this research effort were: (1) a review of the existing literature on the effects of practice and coaching on spatial test scores, (2) a review of manuals, books, and other publications dealing with coaching of spatial test items, (3) the development of specific coaching strategies and materials, and (4) the development of presentation media to teach the coaching strategies to prospective students.

Previous Research on Coaching and Practice Effects on Spatial Tests

An investigation into the previous literature on the effects of practice and/or coaching revealed several research articles which suggest that practice and coaching do, in fact, affect scores on these types of tests. For example, Stericker and LeVesconte (1982) found that spatial ability is significantly affected by prior coaching on how to take the test. Stericker and LeVesconte (1982) also found that visual-spatial skill is trainable in both female and male adults and suggested that the effects of training on certain spatial tests generalized beyond the immediate training situation to increase scores on other tests of spatial ability as well. However, Gagnon (1985) found that a five-hour training session on certain spatial tasks did not generalize to higher scores on other related spatial tests.

In a series of meta-analyses, Baenninger and Newcombe (1989) found that "specific" training (i.e., training on a single spatial measure) produced significant increases in spatial scores. The same authors also found that "short" training (single or brief administrations over a period of less than three weeks) produced effect sizes which were not significantly different from those of practice alone. Baenninger and Newcombe (1989) concluded that "...brief training fulfills the same function as practice. That is, it enhances test-specific spatial ability but not necessarily general spatial ability" (p. 339).

Other researchers have documented similar results. Brinkmann (1966) found that a programmed instruction technique sucessfully improved scores of "spatial visualization". Kyllonen, Lohman, and Snow (1984) noted that training strategies and performance feedback were both effective in increasing the scores on spatial tests. McGee (1978) found that training (in the form of a one-hour lecture on spatial

---

[1] All statements expressed in this paper are those of the authors and do not necessarily reflect the official opinions or policies of ARI, PERSCOM, and the Department of the Army.

abilities) significantly improved scores on a five-item form of the Mental Rotation Test. Saunderson (1973) also found that specific training on spatial tasks was directly related to increased scores on subsequent spatial ability tests. Gibson (1953) reported studies which found that practice and/or training had similar results on improving such spatially oriented skills as estimating the linear extent, area, and angles of various geometric figures. Sherman (1974) found similar results of practice effects on various spatial tests. Goldstein and Chance (1965) and Conner, Schackman, and Serbin (1978) also found substantial training and practice effects on a set of items taken from the Embedded Figures Test and two other measures of field dependence.

In conclusion, a great deal of literature exists which suggests that practice and/or specific coaching on spatial tasks improves scores on spatial ability tests.

Coaching Manuals, Books, and Publications

The second part of this investigation was to survey actual "coaching" books and manuals which are easily accessible to the public (i.e., recruiters and recruitees). Two examples are: Up the I.Q.! by Paul Jacobs (1977) and Know Your Own I.Q. by H. J. Eysenck (1962). Jacobs' book presents a list of 12 principles or rules for solving test items and provides numerous examples and practice items to fascilitate learning. The Army's Figural Reasoning test, one of the measures of concern, contains items which are very similar to those in Jacobs' book. Eysenck's book also includes this type of item and supplies readers with the correct answers and how they can be obtained.

In summary, there appears to be cause for concern about the long-term validity of the Project A spatial tests. Spatial test scores in general seem to be susceptible to coaching and/or practice effects and coaching aids are readily available for this purpose. Therefore, the possibility remains that scores on the Project A spatial tests may be invalidated by the confounding influence of differential coaching and/or practice experience. However, since coaching involves training on specific test items (Anastasi, 1982), the results of previous research are not conclusive and more focused research is necessary.

Developing Coaching Strategies

The next stage was to develop specific coaching strategies for the Project A spatial tests (Assembling Objects, Figural Reasoning, and Orientation). In general, there were several steps taken to develop strategies, clues, and "hints" designed to make the spatial items quicker and easier to solve.

Over all test items, the first step in the development of strategies involved breaking the tests down and grouping similar items. In general, the analysis of similar item types was based on patterns of problem solutions, object shapes, number of connections, and amount of rotations. The second step was to count the similarities in types of items to determine the strategy that would answer the largest number of items. In the third step, strategies or "hints" were developed to represent the most common similarities to the least common. Hence, the most important criteria used in developing the coaching strategies was to choose the similarities in items that helped to solve the largest number of items in the test.

The first test, Assembling Objects, contains 36 items with an 18-minute time limit. The subject must choose from four possible answers. The first part of this test involves connecting pieces together that are labeled with identical letters to "build" or "assemble" an object. Figure 1 is a fictitious item (similar to the fictitious items used in the coaching strategy) designed to represent an Assembling Objects item.

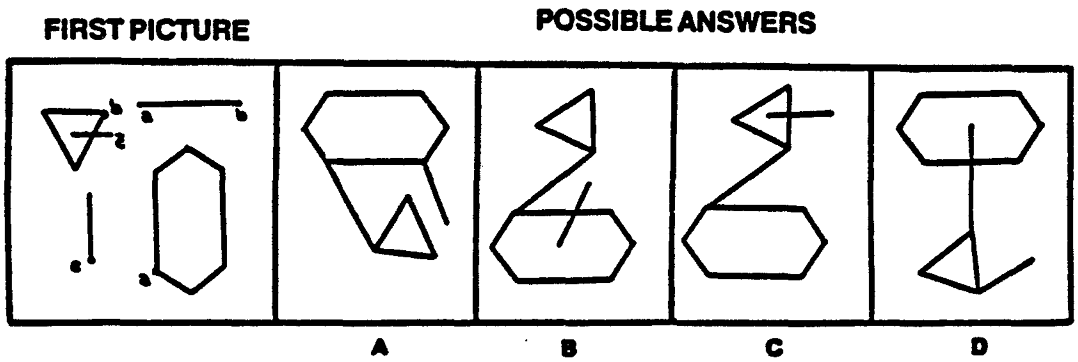


Figure 1. Example of Assembling Objects Item (Part 1)

The coaching strategy for this type of item was to teach the subjects to eliminate wrong answers by simply looking at the number and location of connections for each item. On many of the items, it was found that the number of connections on a given object could eliminate one , two or all three of the distractors without ever having to "assemble" the pieces to arrive at the correct solution. It was determined that this simple trick helped eliminate more distractors (and solve more items) than any other generalities or similarities found in this test.

The second part of the Assembling Objects test involved fitting object shapes together like a puzzle into one larger object shape (Figure 2).

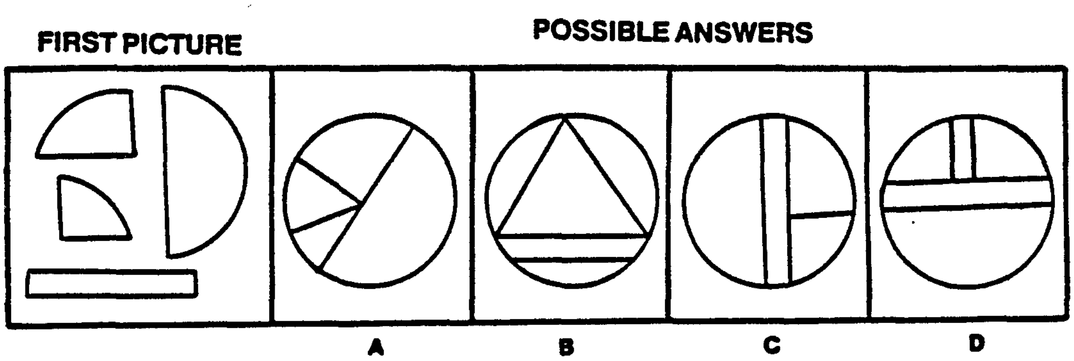


Figure 2. Example of Assembling Objects Item (Part 2)

The coaching strategy for this section was to teach subjects to eliminate answers that didn't contain the largest object shape, the smallest object shape, and the same number of object shapes as in the first picture. The largest object shape was the main focus of the coaching strategy because, again, it was responsible for solving the largest number of items in this part of the test. In other words, eliminating those distractors that did not have the largest object shape included solved the largest number of items on the test by itself without having to piece the object shapes together.

The second test, Figural Reasoning, consists of 30 items with a 12-minute time limit. The task for this test is to identify the pattern or relationship that exists among a series of four figures and then to identify from five possible answers the figure that should appear next in the pattern/relationship (Figure 3). The coaching strategy developed for this test reviewed in depth the six patterns contained in the test (addition, subtraction, rotation, movement, repetition, and number) and gave subjects instructions and practice in detecting them and choosing the correct answer. For example, the coaching strategy for the repetition pattern teaches the subjects to choose the answer that is identical to the picture presented in the first and third positions in the pattern series. By focusing on these pictures, the subject can automatically eliminate all of the distractors and choose the correct solution to the problem.

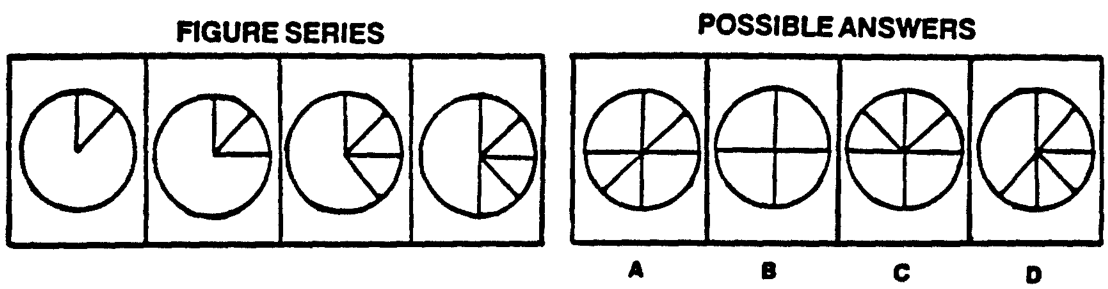


Figure 3. Example of Figural Reasoning Item

The remaining test, Orientation, contains 24 items with a 10-minute time limit. Each item contains a picture within a circular or rectangular frame (Figure 4).

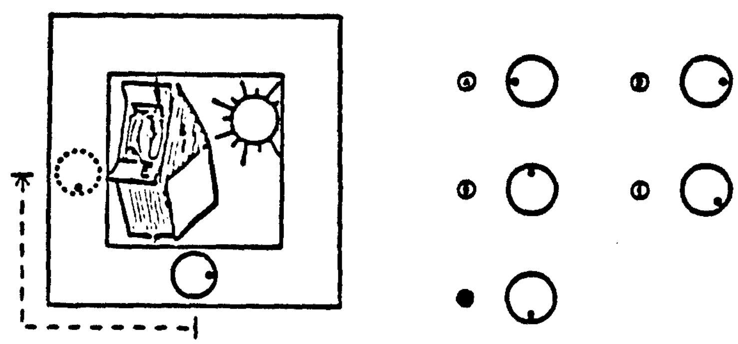


Figure 4. Example of Orientation Item

The bottom of the frame has a small circle with a dot inside it. The picture or scene is not in an upright position and the task is to mentally rotate the frame so that the small circle with the dot is positioned at

the bottom of the picture. After doing so, the subject must then decide where the dot will appear in the small circle and choose the correct answer from among five alternative answers. The coaching strategy for this test involved telling subjects that all rotations are in multiples of 1/8 turns. Subjects are then instructed and given practice on how to count the number of turns necessary to align the small circle and dot with the bottom of the picture. Counting off the same number of 1/8 turns in the same direction within the small circle with the dot gives the correct answer. This coaching strategy was seen by the authors as the most effective because it effectively reduced the spatial orientation exercise into one of a mathematical (counting) exercise.

The Coaching Presentation Media

The coaching presentation medium which was used in the final research design evolved from numereous constraints on the overall experiment. The first design was to present the coaching material by means of overhead projection slides and a live, oral presentation of the strategies. However, due to lack of classroom space and poor angles of vision, the overhead projection was ruled out in a preliminary pilot test. The idea of a videotape presentation, prefered for ease and consistency in administration, was ruled out due to the small image size of the item examples on the television monitor which would drastically reduce the size of the test groups able to view the videotape. The final presentation medium tested and finally used in the research experiment involved an audiotape presentation of the strategies accompanied by a workbook of examples and practice items for the student to use as they listened. This medium solved all visual problems encountered with the earlier ideas and presented the example and practice items in the same physical format as in the test itself. Due to this, the transfer effects of coaching and practice should have been maximized in the experiment (Busciglio, 1992).

References

Anastasi, A. (1982). Psychological testing (fifth edition). New York: MacMillan.

Baenninger, M. & Newcombe, N. (1989). The role of experience in spatial test performance: A meta-analysis. Sex Roles, 20, 327-344.

Brinkman, E.H. (1966). Programmed instruction as a technique improving spatial visualization. Journal of Applied Psychology, 50, 179-184.

Busciglio, H.H. (1992, October). Assessing the coachability of Project A spatial tests: Empirical results. Paper presented at the annual conference of the Military Testing Association, San Diego, CA.

Conner, J.M., Schackman, M., & Serbin, L.A. (1978). Sex-related differences in response to practice on a visual-spatial test and generalization to a related test. Child Development, 49, 24-29.

Eysenck, H.J (1962). Know your own I.Q.. Baltimore: Penguin.

Gagnon, D. (1985). Videogames and spatial skills: An exploratory study. Educational Communication and Technology, 33(4), 263-275.

Gibson, E.J. (1953). Improvement in perceptual judgements as a function of controlled practice or training. Psychological Bulletin, 50, 401-431.

Goldstein, A.G., & Chance, J.E. (1965). Effects of practice on sex-related differences in performance on Embedded Figures. Psychonomic Science, 3, 361-362.

Jacobs, P.I. (1977). Up the I.Q.! New York: Wyden.

Kyllonen, P.C., Lohman, D.F., & Snow, R.E. (1984). Effects of aptitudes, strategy training, and task facets on spatial task performance. Journal of Educational Psychology, 76, 130-145.

McGee, M.G. (1978). Effects of training and practice on sex differences in mental rotation test scores. Journal of Psychology, 100, 87-90.

Saunderson, A. (1973). The effect of a special training programme on spatial ability test performance. New Guinea Psychologist, 5, 15-23.

Sherman, J.A. (1974). Field articulation, sex, spatial visualization, dependency, practice, laterality of the brain and birth order. Perceptual and Motor Skills, 38, 1223-1235.

Stericker, A., & LeVesconte, S. (1982). Effect of brief training on sex-related differences in visual-spatial skill. Journal of Personality and Social Psychology, 43, 1018-1029.

ASSESSING THE COACHABILIITY OF PROJECT A SPATIAL TESTS: EMPIRICAL RESULTS[1]

Henry H. Busciglio
U. S. Army Research Institute

Introduction

Army researchers developed specific coaching strategies to determine the extent to which examinees' scores on three Project A spatial tests can be inflated through coaching (Palmer & Busciglio, 1992). These strategies and materials were then used in an experiment that also assessed the vulnerability of the tests to the effects of practice and more general coaching.

Method

Development of General Coaching Strategy

To assess the degree to which the tests were susceptible to more traditional types of multiple-choice coaching, we scanned a number of popular coaching references (e.g., Barron's Educational Series, 1989; C.E.E.B., 1983; Steinberg, 1987) to develop a single-page handout listing hints on "Doing Better on Multiple-Choice Tests." These hints included such things as time management and guessing strategies. All subjects assigned to one of the general coaching conditions (see below) received the same handout, regardless of the test taken.

Subjects

Data were collected from 1,915 receptees at Fort Jackson, South Carolina in June of 1992. All subjects were tested in two-hour sessions, in groups of approximately 40 to 120 persons. In general, groups receiving specific coaching were smaller, while groups receiving general coaching or practice only were larger.

Testing Schedules

Subjects were divided into one of fifteen groups. The second column of Table 1 summarizes the testing schedules used. Subjects assigned to Groups 01, 06, and 11 took one of the three tests, then listened to the audio tape and studied the workbook containing the specific coaching strategy for the test, then retook the test. In Groups 02, 07, and 12, subjects received specific coaching before taking the test for the first time; then, after a short break, subjects retook the test. Subjects in Groups 03, 08, and 13 took one of the three tests, then received the handout giving coaching on general test-taking strategies

[1]Presented at the annual meeting of the Military Testing Association, 27 Oct to 29 Oct, 1992. All statements expressed in this paper are those of the author and do not necessarily reflect the official opinions or policies of the U.S. Army Research Institute or the Department of the Army.

Table 1

Descriptive Statistics and Effect Sizes

Group	Schedule[a]	Sex	N	1st Test M	1st Test SD	2nd Test M	2nd Test SD	Effect Size[b]	t[c]
01	O AO O	M	117	23.2	6.9	25.0	7.8	0.27	2.958**
		F	50	20.9	6.7	24.0	6.9	0.46	3.845***
		All	167	22.5	6.9	24.7	7.5	0.32	4.434***
02	AO O O	M	51	25.5	7.0	28.1	6.6	0.37	2.951**
		F	104	24.7	6.7	28.1	5.9	0.51	6.206***
		All	155	24.9	6.8	28.1	6.1	0.46	6.721***
03	O GN O	M	108	19.1	8.4	21.9	8.8	0.33	3.648***
04	GN O O	M	117	20.7	7.2	23.5	8.3	0.38	5.390***
05	O O	M	120	18.2	7.5	23.1	8.4	0.66	8.594***
06	O FR O	M	105	20.3	4.9	23.2	4.4	0.60	7.455***
		F	58	20.2	5.5	23.1	4.2	0.53	6.566***
		All	163	20.3	5.1	23.2	4.3	0.57	9.817***
07	FR O O	M	171	21.2	5.5	21.4	6.4	0.04	0.712
08	O GN O	F	96	18.0	6.2	19.9	6.0	0.31	5.643***
09	GN O O	M	110	20.2	4.5	21.3	5.3	0.24	2.881**
10	O O	F	60	19.6	5.5	22.3	4.9	0.49	5.625***
11	O OR O	M	222	10.8	6.2	16.3	6.7	0.90	13.657***
12	OR O O	M	116	15.5	7.6	15.8	8.0	0.17	3.345**
		F	36	11.7	7.1	13.3	7.7	0.23	2.750**
		All	152	13.9	7.6	15.2	8.0	0.18	4.209***
13	O GN O	F	111	8.7	5.5	10.3	6.0	0.29	4.126***
14	GN O O	M	59	10.8	5.8	12.7	7.6	0.32	3.250**
15	O O	M	55	11.3	6.4	12.6	7.4	0.21	2.248*
		F	47	8.7	5.3	9.1	5.8	0.08	0.680
		All	102	10.1	6.0	11.0	6.9	0.15	2.135*

[a]O = Testing; GN = General Coaching; AO, FR, OR = Specific coaching on Assembling Objects, Figural Reasoning, and Orientation tests, respectively. [b] Effect size = (2nd Test mean - 1st Test mean)/SD on 1st Test. [c]t is for dependent groups test: ***p<.001 **p<.01 *p<.05.

before retaking the test; all subjects, regardless of the test taken, received the same handout. In Groups 04, 09, and 14, subjects received general coaching before taking the test for the first time; then, after a short break, subjects retook the test. Finally, subjects assigned to Groups 05, 10, and 15 took one of the three tests, then had a short break before re-taking the same test.

Results

Descriptive Statistics and Effect Sizes

Table 1 shows descriptive statistics and effect sizes for all subjects in the experiment. As can be seen, mean scores on the second testing in almost all groups were significantly higher than means on the first testing. Especially noteworthy are the large effect sizes for practice on the Assembling Objects test (0.66 for Group 05) and specific coaching on the Orientation test (0.90 for Group 11). Equally noteworthy is the very small practice effect for the Orientation test (0.15 for all subjects in Group 15).

Along with the within-subjects results reported in Table 1, a number of between-subjects effects were assessed. These are reported separately below.

Effects of Coaching on Pretest Scores

The first analysis assessed the extent to which groups receiving specific or general coaching prior to their first testing (e.g., Groups 02 and 04, respectively) scored significantly higher on the first testing than did those groups receiving no coaching before the first testing (e.g., Groups 01, 03, and 05). The top portion of Table 2 shows the results of the ANOVA and subsequent comparison of cell means. On all three tests, groups receiving specific coaching achieved significantly higher scores than did those groups receiving no coaching. In contrast, groups receiving general coaching did no better, on any of the tests, than did the groups getting no coaching. [Readers should interpret these results with caution; although the assignment of groups to experimental conditions was done in an unbiased manner, the absence of pretest data makes it impossible to know if the groups were indeed equivalent.]

Effects of Coaching on Pretest-Posttest Gain Scores

This analysis assessed the extent to which specific and general coaching (e.g., Groups 01 and 03, respectively) between a pre- and posttest led to gain scores that were significantly greater than those for the groups having only practice (e.g., Group 5) between pre- and posttests. The middle portion of Table 2 shows the results of the ANOVA and subsequent comparison of cell means. As can be seen, the Orientation test was the only measure for which specific coaching produced gain scores

Table 2

Between-Subjects Analysis of Coaching Effects

	Test					
	Assembling Objects		Figural Reasoning		Orientation	
On Pretest Scores:						
ANOVA	F	df	F	df	F	df
	23.22***	2,665	5.97**	2,597	19.60***	2,643
Cell Means	N	Mean	N	Mean	N	Mean
Specific	155	24.95(a)	171	21.21(a)	152	13.88(a)
General	117	20.72(b)	110	20.16(a)(b)	59	10.85(b)
None	396	20.20(b)	319	19.45 (b)	435	10.11(b)
On Pretest-Posttest Gain Scores:						
ANOVA	F	df	F	df	F	df
	5.84**	2,392	2.55	2,316	36.55***	2,432
Cell Means	N	Mean	N	Mean	N	Mean
Specific	167	2.22(b)	163	2.93(a)	222	5.54(a)
General	108	2.76(b)	96	1.89(a)	111	1.59(b)
None	120	4.94(a)	60	2.73(a)	102	0.93(b)
On Gain Scores Between Two "Posttests":						
ANOVA	F	df	F	df	F	df
	4.86**	2,389	9.50***	2,338	0.91	2,310
Cell Means	N	Mean	N	Mean	N	Mean
Specific	155	3.14(b)	171	0.21(b)	152	1.36(a)
General	117	2.74(b)	110	1.09(b)	59	1.85(a)
None	120	4.94(a)	60	2.73(a)	102	0.93(a)

Note. * $p<.05$. ** $p<.01$. *** $p<.0001$. Means marked with the same letter are not significantly different ($p<.05$, Tukey HSD test).

significantly greater than those for practice alone; indeed, for the Assembling Objects test the reverse was true. For all three tests general coaching led to gain scores equal to or smaller than those for practice alone.

Effects of Coaching on Gain Scores Between Two "Posttests"

As discussed above, some groups received either specific or general coaching before being tested twice. The final analysis assessed the extent to which coaching before these two "posttests" (e.g., Groups 02 and 04) led to gain scores that were significantly different from those among groups having no coaching before the two administrations (e.g., Group 05). The bottom portion of Table 2 shows the results of the ANOVA and subsequent comparison of cell means. As can be seen, both forms of coaching led to gain scores equal to or smaller than those for practice alone.

Discussion

The results shown in Tables 1 and 2 indicate that both coaching and practice lead to significant increases in scores on the three Project A spatial tests. However, the results also suggest that, with the exception of the Orientation test, coaching may not lead to score gains that are significantly greater than those to be expected from practice alone.

In preparing for this experiment, researchers at the Army Research Institute attempted to create the best possible specific coaching strategies for each test (Palmer & Busciglio, 1992). We believe, therefore, that the lack of strong coaching effects - above and beyond the effects of mere practice - on the Assembling Objects and Figural Reasoning tests, supports the view that these instruments are not unreasonably susceptible to invalidation through differential coaching experiences.

Although the magnitude of practice effects on the Assembling Objects and Figural Reasoning tests was not trivial, we would point out that such effect sizes are commonly found on spatial tests (cf. Palmer & Busciglio, 1992) and that our testing schedule can be thought of as a "worst case" example of practice, involving test-retest intervals that are much shorter than can be expected in an operational environment. In any event, it may be possible to control for differential practice effects by giving all examinees the same amount of practice directly prior to the test; that is, administer a certain number of [nonscored] practice items before the scored items.

In the present research, the Orientation test was unique in that specific coaching effects were large and significantly greater than effects for practice alone, indicating that the "trick" involved in the coaching may have given subjects a completely new [nonspatial] strategy, and not simply more and/or better practice (as may have been the case with the Assembling

Objects or Figural Reasoning tests). A number of countermeasures to decrease the usefulness of this type of coaching may be explored in the future, such as rotating the picture in increments other than multiples of 1/8 of a turn (e.g., 1/5, 1/6, 1/7 of a turn).

For all three tests the general coaching was ineffective and may have been counterproductive, producing effect sizes less than those of practice alone. A number of explanations for this finding are possible: 1) nonspatial coaching may simply be inappropriate for spatial tests; 2) hints about guessing may have led subjects to spend less time attempting to work items before guessing.

The assessment of the impact of prior coaching on practice effects between two posttests was meant to gauge the merits of a number of possible hypotheses. Would prior coaching enhance practice effects, perhaps by more narrowly focusing subjects' attention on what should be practiced? Would practice and coaching effects instead be unrelated (and thus additive)? Our results seem to support yet another possibility: that prior coaching decreases the effects of practice because there is only a certain amount of gain possible for each subject before a ceiling is reached.

It only remains to say that preliminary item level analyses have so far given no indication that any coaching or practice effects differed appreciably across items, and that self-report posttest data were gathered on, among other things, the extent to which subjects used the specific coaching strategies. Future analyses of these data may help to better explain the results summarized here.

References

Barron's Educational Series (1989). How to prepare for the Armed Forces Test (3rd. Ed. pp. 7-9). New York: Barron's Educational Series, Inc.

College Entrance Examination Board (1983). 10 SATs: Scholastic aptitude tests of the college board (p. 11). New York: College Entrance Examination Board.

Palmer, D.R., & Busciglio, H.H. (1992, October). Assessing the coachability of Project A spatial tests: Developing strategies. Paper presented at the annual conference of the Military Testing Association, San Diego, CA.

Steinberg, E.P. (1987). Scoring high on the Armed Forces test (pp. 27-30). New York: Pocket Books.

A Principal Components Analysis of 59 Variables Descriptive of Uncovered Spies

LeRoy A. Stone, Ph.D., (Forensic) ABFP, ABPP
Harpers Ferry, WV

In 1987, this investigator started a database for all those U.S. citizens who, since the end of World War II, have been uncovered spying on their own country. From the very beginning, it was decided that only quantifiable descriptive variables would be employed in constructing this database. The primary reason for wanting to build such a database was due to the fact this investigator discovered that no such database existed in the private sector and it also appeared then that such a quantified database (which could be subject to statistical analyses) did not seem to exist, even within governmental agencies and bureaus. To anyone not overly familiar with research conducted on the general topic of uncovered spies, it comes as a shock to learn that no one has apparently produced any systematic scientific research on this subject that has been published or even distributed on a limited basis. There have been a number of publications which accomplished nothing more than providing narrative-type descriptions of a limited segment of the known uncovered spies; the best previous attempts at quantification have occurred only in a couple of publications (e.g., Crawford, 1988; Jepson, 1987). In these, the only resulting numerical analyses were entirely limited to a very few frequency-count tables of matters such as amount of education, number of years of federal or defense-associated employment, age at time when caught, etc.). Not even simple descriptive statistics were used in these publications. In contrast, the present investigator has completed more than several investigations that have employed complex statistical analyses, some of which utilized so-called multivariate mathematico-statistical procedures (e.g., Stone, 1991a, 1991b, 1992a, and 1992b).

When the presently reported research was begun, the U.S. spy database involved 186 spies, each with quantified listings on up to 68 different variables. The choice of variables was mainly dictated by noting what kind of information was available in books, magazine articles, newspaper accounts, Government published security training monographs and the like, information obtained from Federal agencies/bureaus based upon Freedom of Information Act requests, and information derived from commercially available journalistic databases. Some of the variables in this database were psychological measurement estimations (e.g., intelligence quotients) based on estimation procedures which seem to be well-established in certain applied areas of psychology (Wilson, Rosenbaum, Brown, Rourke & Whitman, 1978; Stone, 1991c). The database is one which is certainly not static; it is frequently being added to, either with new variables or more often with the missing data-points being replaced with recently obtained information. It is most unlikely that this particular database will ever be considered as being final and complete. A number of the investigations, which have involved utilization of this database, have been reported in a review article (Stone, 1991a).

Since the original decision, as to whether to include a variable

in this database, has almost always been more determined by the matter of information availability rather than anything else, it becomes particularly important to understand what basic aspects of spies are really being represented by the variables in this database. High redundancy was known to be a problem with a few of the involved variables. It was wondered whether factor analytic or principal components methodology might serve as a means to better understand the particular composition of variables in the spy database. It was especially wondered as to how many principal components would be required to be able to account for given percentages of the common variance for the measurement variables included in the database. Would the most major components be interpretable? Would there be some factors/components that were primarily represented by only a single variable? Could factor/component scores be used to replace some of the redundant variables' measurements? Basically, since this particular database was the first-known, fully-quantified one constructed to be descriptive of caught or uncovered spies, a factor or principal components analysis of same was the focus of a good deal of curiosity as no such previous accomplished analyses of this kind of content have been reported elsewhere.

Method and Results

The particular correlational matrix that was to be submitted to a principal components analysis was originally 68 x 68; however, it was decided, due to a rather major incomplete data measurements problem, to actually analyze 59 of the involved variables. The nine variables, not included in the analysis, were omitted as the numbers of measurement observations on each were less than the total number (i.e., 68) of variables in the database. For most of the analyzed 59 variables, the number of measurements involved with each variable generally differed somewhat, but most were fairly close to the total number of spies described in the database (i.e., 186). Intercorrelations of the analyzed correlational matrix ranged from -0.99 to 0.98. The employed computerized principal components routine made use of a variation of the rather old Jacobi algorithm to obtain trial eigenvalues. The number of principal components extracted and retained was set to be limited to the number of eigenvalues that had values in excess of unity. The resulting number of principal components thusly obtained was 17. It is interesting to note that use of Cattell's scree test produced a graph that suggested consideration of only a five-factor/components limitation. However, for subsequent rotation of the components, this particular limited number of factors or components was not used. The 17 principal components were rotated using Kaiser's varimax normalized procedure (the rotated loadings for these principal components are given in Table 1; the variables have been sorted into order according to principal component loadings sizes). When number of variables is relatively large (i.e., larger than 40 or so), results obtained using a principal components solution usually can be expected to closely resemble those obtained based on most of the other well-known orthogonal factor analysis solutions.

Inspection of Table 1 reveals that the 17 rotated principal components totally accounted for about 81% of the common variance for the involved 59 variables. If one limited the number of principal components to only the first five, the number suggested by employment

of scree graphing, these five are seen to account for only about 40.4% of the common variance, or about half of that accounted for by the full 17 principal components. The first principal component, the largest one, accounted for about 13% of the total common variance and about 16% of the common variance accounted for by the full number of 17 principal components. The communality values range from a low of 0.586 to 0.990. The mean communality value was 0.780.

Interpretation of the Principal Components

Since the present investigator is familiar with what the variables represent, it was not difficult to suggest a label for the first principal component, which was clearly bipolar in form, and this was "Monetary vs. Ideology Motivations for Spying." In previous reports, some of the database variables, such as Foreign preference, Jewish background, Year in which spying started, Birthyear, Having of money problems, etc., have already been noted (e.g., Stone, 1991b, 1992a) to be highly correlated with a similar derived motivational variable. The second principal component can be rather easily labeled as a sort of "Mental Ability" construct. The third component looks to be an "Age/Experience" matter. The fourth component looks like a bipolar "Military vs. Civilian Background" construct. A label of "Nativeborn Origins" seems appropriate for the fifth component. "Substance Abuse" represents a very clear title for the sixth component. The seventh component seems to justify a name of "Length of Sentence Given for Spying". "Sexual Misconduct" is a likely title for the eighth component. The ninth component is not really easy to name; however it, when some interpretation is forced, looks a bit like "Air Force Spies (but not Navy Spies) Described in Government-Produced Publications." The tenth component is also not easily interpreted; one can impose a label of something such as: "Army Spies Arrested Overseas by a Foreign Power Agency." The eleventh component seems easy to understand; it rather clearly can be defined as a "Mental and Emotional Problems Involving Criminal or Deceptive Actions" kind of construct. The twelveth component can be interpreted as a sort of "Coerced into Spying (and Not Having Financial Problems) to Benefit Countries Other Than the Soviet Union" matter. "Long-term Spying Ending in Suicide" seems to be an appropriate name for the thirteenth component. Component fourteen can be understood as a sort of "Disaffection Motivation as the Basis for Spying (and Being in the Army)" matter. Component fifteen can be labeled as a kind of "Criminally Based Volunteering to Spy (More Frequently Found in the Air Force)" factor. "Intelligence Community Background Associated with Previous Security Violations" can be used to describe the sixteenth component. The seventeenth and last extracted component can be labeled as a "Black-American Marines Overseas Location" matter.

It is encouraging to note that most of these extracted and rotated orthogonal principal components can be somewhat readily named and seem to be perhaps interpretable. Also, noted is that each of these components has more than one of the variables loading high on them; none of the components seems to be just a single-variable kind of factor. Interpretability seems not to decline much even after the fifth component (which, according to scree charting, seemed to suggest that it be considered as the final one).

Table 1
Varimax Rotated Principal Components Matrix

Variables	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	h2
	Principal Components*																	
57**	94*	-17	02	15	05	03	08	-01	-07	01	12	06	06	-04	07	-03	01	975
17	-90*	17	-02	-10	-03	-10	-06	-01	08	-03	-11	08	01	-04	-02	09	-01	896
59	81*	-15	-11	06	-01	13	04	-07	-10	03	00	02	21	37	13	-03	-02	919
16	81*	-08	00	15	08	-10	15	-07	-07	-04	13	31	09	-27	07	07	04	930
10	79*	-01	-03	-03	-17	18	-06	-01	29	-05	-23	-07	-09	00	02	12	17	897
29	-76*	10	-07	-23	-15	-04	04	-11	32	-03	03	05	17	-12	22	-08	06	891
39	-73*	12	06	-04	-09	-07	04	-22	08	-08	-11	17	-12	-06	-10	-12	13	734
56	71*	06	-15	04	06	18	-17	-16	30	04	-21	-11	-35	01	00	16	10	927
8	64*	-22	-54*	07	-05	13	-08	-06	28	01	-17	01	-17	06	06	08	14	949
49	62*	01	11	02	22	03	-27	-20	02	25	00	16	00	-33	-09	-11	-06	786
51	46*	-21	29	12	-14	07	16	-26	-27	-05	-02	-01	-27	11	-06	12	28	723
13	-15	94*	16	-14	-05	-05	00	-01	-02	-06	-04	01	03	-07	-02	04	-12	990
14	-15	94*	16	-14	-05	-05	00	-01	-02	-06	-04	01	03	-07	-02	04	-12	990
12	-13	93*	24	-15	-08	-06	-04	-02	01	-01	-03	-04	07	-05	-05	04	-02	984
9	-21	83*	12	-19	-08	-08	-12	-02	06	13	-05	-17	13	03	-03	00	10	892
23	-09	61*	52*	01	01	-16	05	-04	-04	08	06	05	05	07	-17	-10	10	753
1	-13	51*	-23	01	11	10	09	00	-07	-29	01	22	-28	-03	08	03	-42*	760
20	12	14	85*	21	-07	07	10	03	-03	06	-04	-05	15	04	03	13	03	873
22	04	28	83*	-09	07	01	-05	-03	-10	05	07	-10	-03	03	-10	06	-05	832
11	-12	36	79*	-23	-09	00	02	07	-02	-09	02	-07	21	02	-11	-04	-05	915
43	-17	00	65*	-17	-03	-03	06	-03	15	-18	-17	08	-01	-05	14	-30	-12	702
28	33	-31	-18	75*	04	04	04	20	01	-08	00	10	00	03	17	03	00	899
3	-33	26	23	-75*	-04	00	-05	-22	03	10	03	-09	-02	01	-19	-01	03	906
54	05	-38	04	73*	-11	-04	-13	-18	-05	13	01	02	-19	-14	-01	-11	15	852
44	23	23	18	50*	13	-04	04	12	-11	36*	-22	-17	19	33	10	14	-10	832
40	44*	-13	06	48*	08	21	03	-17	20	-06	-04	-06	-04	-03	-22	05	-03	619
2	31*	30*	-01	42*	01	-05	19	-15	-08	25	19	31*	04	27	03	-18	29	813
24	06	-10	-01	03	93*	06	02	03	05	03	-04	01	-02	-01	05	03	02	903
41	10	-11	01	03	91*	10	02	01	02	02	-04	-01	-07	-05	03	05	05	885
52	07	24	-16	-27	50*	-06	26	28	11	10	-20	07	18	-10	14	-07	-05	693
58	21	-11	-05	02	09	91*	09	06	04	07	06	01	06	06	04	05	-02	926
35	12	-04	15	02	08	88*	06	15	-04	01	05	-05	08	10	09	-13	00	893
36	21	-13	-22	04	00	59*	12	-19	28	14	-01	13	-07	17	-08	29	00	756
15	11	01	08	04	02	18	91*	-03	04	02	-04	00	00	00	-06	06	-07	889
25	-26	-17	03	-01	06	06	86*	-14	-16	07	04	08	-13	-13	18	-16	06	990
47	25	16	04	-52*	16	-13	53*	03	-17	15	-10	12	-15	00	04	16	01	806
38	00	-19	02	-07	00	11	-11	82*	-02	02	-06	-05	01	-02	-03	14	19	804
27	-06	15	01	31	17	07	-04	59*	05	20	09	08	-20	00	-14	07	-07	633
33	-02	08	16	-12	-30*	-19	02	34*	07	-27*	08	-21	21	27*	24	00	29*	675
50	17	01	09	01	-08	-13	26	-04	-72*	13	-13	-06	-03	-05	-02	23	08	747
5	23	-05	-24	44*	08	08	09	-05	53*	18	-14	04	-03	-18	-10	13	-21	767
42	-14	04	24	-01	02	-08	18	30	44*	-06	14	-26	-17	21	-10	-07	32*	680
6	10	-04	-07	25	-33*	12	-07	26	-42*	17	15	08	-22	-14	39*	-17	-10	770
48	-07	11	00	-03	-06	-07	-10	-12	09	-77*	01	04	-07	01	-09	-01	11	682
4	10	-22	14	33*	25	-23	06	-02	-19	-44*	01	-01	26	41*	-05	-07	-18	811
55	02	26	-20	-37*	21	20	08	-18	06	-37*	-16	06	04	-12	-22	12	-07	619
32	03	06	-08	10	-12	04	-03	10	-07	-12	84*	-10	-02	-13	-06	-06	-09	812
37	02	-20	12	-21	-06	03	01	-09	21	16	69*	11	-09	20	06	02	07	721
31	21	-08	-06	13	07	21	-09	-27	01	-23	41*	07	-02	-01	39*	00	-03	586
19	04	-02	-02	-12	-01	12	-05	19	-09	11	-04	-75*	-08	01	-07	-28	-07	750
46	-09	-15	-24	-04	00	20	03	23	-06	02	-04	69*	-01	12	-10	-26	-09	769
34	37*	-05	37*	-19	10	00	04	10	-12	19	-10	43*	-14	-10	32*	-08	-11	729
26	04	14	17	-1	00	13	-15	-13	-04	12	-05	02	69*	-02	01	02	05	600
21	-29	20	17	-30	-26	-02	10	15	19	-26	-17	09	52*	-07	-05	-16	02	792
18	03	-10	01	-05	-10	25	-12	-02	07	-03	00	07	-08	74*	-04	00	-03	662
45	02	-11	-07	08	10	04	09	-08	-02	07	-02	01	02	-01	82*	02	01	736
30	06	02	-05	-07	07	01	-02	12	-10	-02	-02	05	-07	-06	-05	80*	14	708
53	11	17	11	08	01	05	03	12	01	28	-14	-03	29	29	30	52*	-21	745
7	06	-18	-21	03	11	01	-03	18	-08	-17	-09	04	03	-09	00	16	78*	817

*Decimals omitted to conserve space

*Loadings in row that are the row s highest plus any which has squared value greater than 1/2 of the square of the largest loading

**The original order of variables was changed so as to better portray the groupings of higher loadings on each of the rotated principal components. Identification of the variables' numerical coding is as follows: 1 = Race; 2 = Gender; 3 = Civilian; 4 = Army; 5 = Navy; 6 = Air Force; 7 = Marines; 8 = Birthyear; 9 = Number of years of education; 10 = Year identified as a spy; 11 = Age when identified; 12 = Estimated verbal IQ; 13 = Estimated performance IQ; 14 = Estimated full scale IQ; 15 = Years sentenced; 16 = Money motivation; 17 = Ideology motivation; 18 = Disaffection motivation; 19 = Other motivation; 20 = Years of federal service; 21 = Months engaged in spying; 22 = Age when spying began; 23 = Employment level; 24 = Born in U.S.; 25 = Sentence time served; 26 = Suicide; 27 = Homosexual; 28 = Military; 29 = Foreign Preference; 30 = Security Responsibility; 31 = Criminal conduct; 32 = Mental/emotional problems; 33 = Foreign Connections; 34 = Financial matters; 35 = Alcohol abuse; 36 = Drug abuse; 37 = Falsification; 38 = Sex promiscuity; 39 = Jewish; 40 = Have SSN information; 41 = Native burn; 42 = Birth state information; 43 = Married; 44 = Clearance level; 45 = Volunteer; 46 = Soviet Bloc; 47 = Civilian agency arrest; 48 = Foreign agency arrest; 49 = Money problems; 50 = Government publication; 51 = Government agency file information; 52 = Nongovernment publication; 53 = Intelligence Community; 54 = Military agency arrest; 55 = U.S. location; 56 = Year started; 57 = Monetary-ideology; 58 = Substance abuse; 59 = Monetary-ideology prediction score.

Inspection of the 'larger' extracted and rotated principal components seems to support a kind of belief that the studied spy database does seem to rather include measurement of at least three or four different kinds of motivations for spying. The first and most major component or factor was rather easily identified as being representative of a bipolar motivational construct, "Monetary vs. Ideology Motivations for Spying." This particular construct also very clearly emerged in another analysis of the database, when it was seen following cannonical correlational analyses of portions of the database (Stone, 1991b, 1992a, and 1992b). The present analysis seems to suggest that the other two spying motivational variables (Coercion [i.e., Component 12] and Disaffection motivation [i.e., Component 14]) can be regarded as being orthogonal rather than as bipolar representations on a single continuum, that has been suggested in a two-factor motivational theory for spying (Stone, 1992a). Actually, when developing this particular two-factor theory, I was quite aware that the Other (which includes 'coercion') and the Disaffection motivations were somewhat separate and that forcing them into being bipolar ends of a single dimension was mainly for the purpose of being able to describe a less complex motivational theory. It should be understood that the actual number of spies whose primary motivation for spying was Other or Disaffection was quite small; a very great majority were known to have spied because of greed or ideology.

The ninth principal component, although somewhat difficult to identify or interpret, does have some interesting connotations. It seemingly represents a component involving spies from one particular military service (and excluding another particular military service) whose cases tend to be described in publications produced by the Federal Government. It is interesting to contemplate what possible governmental policies such a component <u>might</u> represent. A single underlying basis for this component might simply be the government-published book, authored by Crawford (1988). Familiarity with the building of this database does cause its developer to believe that the U.S. Navy historically has been prone to only provide a very minimum of information regarding its members who have been uncovered as spies. In fact, it appears that it has been Navy policy to provide no more information to the public, if any at all, than it might be required to do so.

Based upon these principal components, component scores have been computed for each of the 186 spies in the database. These 17 new sets of scores have been used to increase the number of variables in the database to 85. This of course resulted in an increased redundancy of measurement within the database.

The principal components analysis, that has been accomplished with the caught U.S. spy database, has provided some new insights and understandings regarding the database itself. In the future, when new measurement variables are to be considered for adding to those already employed, the potential new variables can be tested to discover whether they might be adding new kinds of information or whether they might be adding additional measures of some of those constructs for which measurement already exists in the database. However, what is viewed as being the single most important matter in the accomplishment of the principal components analysis of the database is simply the

fact that the database possessed sufficient measurement quantification to warrent such an exploratory, but sophisticated statistical, analysis. Of any known existing caught-spy databases, this particular one seems to be the only one that allows for such systematic analyses. This kind of situation brings to mind the comment once made by Lord Kelvin: "Until you have measured it, you don't know what you are talking about."

References

Crawford, D. J. (1988). Volunteers: The betrayal of national defense secrets by Air Force traitors. Washington, DC: U.S. Government Printing Office.

Jepson, L. P., Jr. (1987). Espionage directed towards U.S. military interests since World War II. Washington, DC: Counterintelligence Division, Directorate for Security and Counterintelligence, Defense Intelligence Agency.

Stone, L. A. (1991a). I spy a myth. Security Management, 35 (10), 26-32.

Stone, L. A. (1991b). The use of demographic and biographic information to predict specific spying motivations. Unpublished manuscript.

Stone, L. A. (1991c). Intelligence levels for U.S. citizens identified as spies. Manuscript submitted for publication.

Stone, L. A. (1992a). A two-factor motivational theory for spying behavior. Unpublished manuscript.

Stone, L. A. (1992b). Canonical-correlation between security-clearance adjudication concerns and later motivational causes for espionage behavior. Forensic Reports, 5, 305-316. Also as an abstract in the (1991) Bulletin of the American Academy of Forensic Psychology, 12, 10.

Stone, L. A. (in press). A study of spies who were caught but never punished. Forensic Reports. Also as an abstract in the (1990) Bulletin of the American Academy of Forensic Psychology, 11, 10.

Wilson, R. S., Rosenbaum, G., Brown, G., Rourke, D., & Whitman D. An index of premorbid intelligence. Journal of Consulting and Clinical Psychology, 46, 1554-1555.

Inquiries regarding this paper may be sent to P.O. Box 395, Harpers Ferry, WV 25425.

It should be noted that at the time this paper was delivered, the spy database had been increased in size to include 197 spies.

TURNOVER INTENTIONS
DURING NAVAL OFFICER TRAINING

Captain J.M. Stouffer

Paper presented at the
Military Testing Association
San Diego, California, 26 October 1992

Reviewed by:	Approved by:
J.P. Bradley	F.P. Wilson
Major	Commander
Research Coordinator	Commanding Officer

Canadian Forces Personnel Applied Research Unit
Suite 600, 4900 Yonge Street
Willowdale, Ontario
M2N 6B7

TURNOVER INTENTIONS
DURING NAVAL OFFICER TRAINING

Captain J.M. Stouffer
Canadian Forces Personnel Applied Research Unit
Willowdale, Ontario, Canada

INTRODUCTION

Background

Voluntary turnover during basic officer training takes two forms:

a. transfer from the military occupation (MOC) the candidate originally enrolled in, to another MOC; and

b. departure from both the MOC and service to return to the civilian sector.

CFPARU developed a plan for investigating the reasons for voluntary withdrawal behaviour during basic officer (BOTC) and military occupation training (Agar & Bradley, 1990). The theoretical rationale underlying this research plan was drawn from the model of reasoned action (Azjen & Fishbein, 1980):

> The best predictor of whether or not an individual will remain with an employer or leave is a measure of the individual's intention to leave or stay. Additionally, the intention to leave or stay will be determined by the individual's attitude with respect to staying or leaving and the social pressure (i.e., from those important to the individual) to do so (Azjen & Fishbein, 1980).

Purpose

The purpose of this research was to determine the extent to which naval officer candidates intend to leave the CF or their MOC and to identify the motivational forces behind these turnover intentions.

The views and opinions expressed in this paper are those of the author and not necessarily those of the Department of National Defence.

Cette publication de l'Unite de recherches psychotechniques des Forces canadiennes sera rendu disponible en francais sur demande.

METHOD

Participant Sample

The participant sample included 307 candidates in various stages of naval officer training.

The Survey Instrument

The Canadian Forces Officer Training Questionnaire (CFOTQ) is a self-report instrument measuring the following:

a. intentions to leave/stay with present MOC (INTMOC);

b. intentions to leave/stay with the CF (INTCF);

c. employment needs;

d. attitude towards current CF career (ATTNOW);

e. attitude towards CF career on completion of MOC training (ATTFUT);

f. attitude towards alternate civilian employment (ATTCIV);

g. subjective norms (i.e., social pressure) with respect to leaving one's present MOC (SNMOC); and

h. subjective norms (i.e., social pressure) with respect to leaving the CF (SNCF).

Measurement of Turnover Intentions

Measures of INTCF and INTMOC were obtained with two items, each measured on a seven-point scale. The two items dealt with the likelihood of leaving and remaining in the CF/MOC, respectively. Intention scores were then computed by subtracting the leaving score from the remaining score. Thus, positive scores indicate intentions to remain, negative scores represent intentions to leave, and scores of zero indicate that the candidate is undecided.

Analyses

CFOTQ data were subjected to a range of analyses for the purpose of determining the extent to which:

a. officer candidates intend to leave their MOC and the CF; and

b. turnover intentions were related to social pressure and attitudes toward current CF employment, future CF employment, and alternate civilian employment.

RESULTS AND DISCUSSION

Turnover Intentions

CF turnover intentions. Of the 307 naval officer candidates that responded to items dealing with leaving and staying in the CF:

a. 9 personnel (3%) reported some intention to leave the CF;

b. 15 personnel (5%) were undecided; and

c. 283 personnel (92%) reported some intention to remain in the CF.

MOC turnover intentions. Of those candidates that responded to items dealing with leaving and staying in their current MOC:

a. 23 personnel (8%) reported some intention to leave their MOC;

b. 24 personnel (8%) were undecided; and

c. 257 personnel (84%) reported some intention to remain in their MOC.

Determinants of Intentions to Leave/Remain in the CF

Table 1 (above the diagonal) shows the interrelations among intentions to leave/remain in the CF, the three measures of attitude (ATTNOW, ATTFUT, and ATTCIV), and SNCF. The data show that SNCF is significantly correlated with INTCF. That is, low SNCF scores (i.e., social pressure to leave the CF) are associated with low INTCF scores (i.e., intentions to leave the CF) and high SNCF scores (i.e., social pressure to remain in CF) are associated with high INTCF scores (i.e., intentions to remain in the CF). Results also indicate that ATTNOW, ATTFUT and SNCF scores are positively related to INTCF and that these variables equally influence candidates' turnover intentions. A significant negative correlation was found between ATTCIV and INTCF. Thus, low ATTCIV scores (i.e., low expectancy of employment needs being met in the civilian sector) are associated with high INTCF scores (i.e., intentions to remain in the CF) and high ATTCIV scores (i.e., high expectancy of need availability in the civilian sector) are associated with low INTCF scores (i.e., intentions to leave the CF). This suggests that the availability of employment needs in the civilian sector may compel candidates to leave the CF, while the absence of valued employment needs in the civilian sector may encourage candidates to remain in the CF.

Table 1

Relations among CF/MOC Turnover Intentions, Attitude, and Subjective Norms

	INT	ATTNOW	ATTFUT	ATTCIV	SN
INT	-	.41	.41	-.16	.41
ATTNOW	.35	-	.87	.23	.23
ATTFUT	.38	.82	-	.26	.23
ATTCIV	-	.23	.26	-	.20
SN	.38	.30	.32	-	-

Note. Only statistically significant correlations are included and all are significant at the .005 level. Correlations above the diagonal relate to intentions, subjective norms, and attitudes toward the CF, while correlations below the diagonal relate to intentions, subjective norms, and attitudes with respect to MOC. INT - intention to leave or remain (i.e., INT above the diagonal = INTCF, and INT below the diagonal = INTMOC); ATTNOW - attitude towards current employment; ATTFUT - attitude towards future employment; ATTCIV - attitude towards civilian employment; SN - subjective norms toward leaving or remaining (i.e., SN above the diagonal = SNCF, and SN below the diagonal = SNMOC).

Determinants of Intentions to Leave/Remain in the MOC

Table 1 (below the diagonal) also depicts the relations among intentions to leave/remain in the MOC, the three measures of attitude (ATTNOW, ATTFUT, and ATTCIV), and SNMOC. Results indicate that SNMOC and ATTFUT equally influence INTMOC. The data show that SNMOC, ATTNOW, and ATTFUT are positively correlated with turnover intentions. That is, low scores on INTMOC (intentions to leave) are associated with low scores on these three variables and high scores on INTMOC (intentions to remain) are associated with high scores on the three variables. A non-significant correlation was found between ATTCIV and INTMOC. Again, although SNMOC and ATTFUT have the strongest relations with turnover intentions, it is likely that ATTNOW also serves to hold some individuals in their present MOC.

Determinants of Turnover Intentions

Relations among subjective norms and attitude. The correlations depicted in Table 1 suggest that candidates' turnover intentions are related to social pressure (i.e., SN) and employment attitudes. A series of stepwise multiple regression analyses were performed to determine which variable was the primary determinant of turnover intentions. The results of these analyses confirmed that SN accounts for the majority of the CF/MOC intention variance. Intentions to leave or remain in the CF/MOC at this point in candidates' careers are most influenced by the social pressure exerted on them by significant others (e.g., family and friends). Given the limited knowledge that candidates have about the CF at this time in their career, it is not surprising that friends, relatives, etc., have a strong influence on their employment and career attitudes. As candidates acquire more personal experience and knowledge about their CF careers, it is possible that their attitude will become the primary determinant in any turnover decision.

Summary

The purpose of this research was to determine the extent to which candidates intend to leave the CF or their MOC and to identify the motivational forces behind these turnover intentions. The results show that employment attitudes with respect to staying or leaving and the social pressure to do so are significantly related to turnover intentions. Of course, future research is required to confirm that turnover intentions are predictive of actual turnover behaviour. For a more detailed account of the research described in this paper, the reader is referred to Stouffer (1992).

REFERENCES

Agar, S., & Bradley, J. P. (1990). A research plan for investigating voluntary attrition during MARS officer training (Technical Note 15/90). Willowdale, Ontario: Canadian Forces Personnel Applied Research Unit.

Azjen, I., & Fishbein, M. (1980). Understanding attitudes and predicting social behavior. Englewood Cliffs, NJ: Prentice-Hall.

Stouffer, J. (1992). Motivational influences on naval officer training attrition. (Technical Note 6/92). Willowdale, Ontario: Canadian Forces Personnel Applied Research Unit.

VERTICAL SLICE BATTALION-LEVEL ASSESSMENT OF FUTURE COMMAND AND CONTROL SYSTEMS

Carl W. Lickteig
Army Research Institute
and
Steven G. Williams and Donald L. Smart
Micro Analysis and Design

Abstract

To more efficiently assess future command and control performance at the battalion level, a series of Data Collection Exercises (DCEs) were developed. This paper documents the design of these DCEs as sequential mission segments taken from a full-scale offensive mission. The exercises are intended to provide reliable, cost-effective data on soldier-in-the-loop, command and control performance by manning selected duty positions at each battalion echelon. Assessment issues for the DCEs are based on the functional requirements for tactical Battlefield Operating Systems (BOS), and each exercise requires participants to execute one or more tasks from the battalion-level Mission Training Plan (MTP).

The DCEs assume a Distributed Interactive Simulation (DIS) environment such as the Close Combat Test Bed (CCTB) at Fort Knox with tank simulators equipped with a Combat Vehicle Command and Control (CVCC) system. To reduce troop and equipment costs, the DCEs employ selected DIS technologies including: teleportation of simulators to standardize battlefield conditions at the start of each exercise, semiautomated forces to "round out" the friendly vertical slice and control enemy force activities, and automated battlefield reports by semiautomated friendly units networked and tethered to the manned simulators of exercise participants.

Introduction

The U.S. Army's current focus on precision warfare, the nonlinear battlefield, and multinational contingency operations only underscores its long-standing requirement for advanced command and control (C^2) capabilities (Foss, 1991). Vehicle-based C^2 systems, however, may deluge tactical commanders in a flood of information if candidate systems are not rigorously developed and tested to meet the unique and pressing requirements of front-line commanders (Giboney, 1991). To identify user-based system and training requirements, the Army Research Institute (ARI) at Fort Knox participates in a bilateral, research and development program on future Combat Vehicle Command and Control (CVCC) systems sponsored by the Tank Automotive Command (TACOM).

The CVCC system is an integrated set of components designed initially for the tank weapon system. The primary CVCC component is a command and control display in the tank commander's station that graphically depicts a tactical map of his operational area and updates the battlefield situation based on digital report and overlay transmissions. This map also displays the real-time locations of all CVCC-equipped friendly units.

CVCC-type systems are expected to support the Army's interoperability on the future battlefield by providing battalion-and-below vehicle commanders an integrated weapon system that enables decisive maneuver, hunter-killer engagement, and the coordination of direct and indirect fire. Within and across units, CVCC's digital network should enhance the ability

of front-line commanders to synchronize plans, coordinate mission preparation, and directly monitor mission execution.

In the CCTB, ARI-Knox has conducted an incremental series--from individual tank to battalion--of soldier-in-the-loop evaluations on future C^2 system configurations. This work includes CVCC company (Leibrecht et al. 1992) and battalion evaluations and the Data Collection Exercises described herein. The CVCC research program is designed to exploit the simulation-based technologies available in the Armor Center's Close Combat Test Bed (CCTB).

CCTB Technologies

The CCTB's Distributed-Interactive Simulation (DIS) technology links developmental and conventional weapon system simulators using local- and long-haul digital networks. DIS battlefield dynamics elicit a collective level of perceptual realism in the simulation of meaningful combat operations (Alluisi, 1991). In the developmental test bed of the CCTB, eight CVCC-configured tank simulators emulate the functions, capabilities, and soldier-machine interfaces anticipated for vehicle-based automated C^2 systems.

With only eight CVCC-equipped tank simulators, additional CCTB technologies are required to conduct company and battalion evaluations. The CCTB's semiautomated forces, called BLUFOR, are used to "round out" the battalion unit and semiautomated OPFOR provide the opposing force. Semiautomated BLUFOR simulators and units are controlled by computer operators under the real-time direction of their unit commanders in the manned simulators. A technology called tethering is used to yoke semiautomated BLUFOR units to the movements of manned simulators. Tethering better ensures that the battalion's C^2 performance and overall mission execution are dependent on participants' performance.

On the simulated battlefield, manned and semiautomated simulators can be teleported at the start of each exercise to prespecifed battlefield locations on the CCTB's digital terrain data base. Teleportation provides an effective tool for both standardizing battlefield conditions for each test unit and generalizing the assessment over differing battlefield locations and situations. It is also an efficient technique for rapidly executing multiple exercises to reduce troop support and other evaluation resource requirements.

A key CCTB technology used in the CVCC research program is automated battlefield reporting by the semiautomated BLUFOR units. Appropriate to their operational situation, semiautomated BLUFOR generate real-time battlefield communications such as Contact, Spot, Shell, and Situation reports. For conventionally equipped units, the BLUFOR operators relay these reports to their participant unit commanders over voice radio. For CVCC equipped units, these BLUFOR messages and BLUFOR simulator locations are

automatically relayed to the command and control display of their unit commanders. Automated battlefield reporting standardizes communications from unmanned elements while generating the information management requirements associated with fully manned units during mission execution.

Vertical and Horizontal Evaluation Slices

For battalion-level assessment with a limited number of manned simulators, the CVCC evaluation complements a full-mission test scenario with the Data Collection Exercises (DCEs). While the DCE's are the focus of this paper, their merit may be best determined as counterpart in a balanced design.

For the full-mission defensive scenario, test participants are assigned to duty positions in company and battalion command to carve a horizontal slice of the battalion's primary command structure. Measures of performance target the horizontal flow of information between company and battalion commanders, and the operational effectiveness of the battalion with respect to mission requirements. The battalion's remaining 48 tanks are BLUFOR units operating under the direction of their sim-based participant commanders.

For the DCE's, troop assignments result in a fully manned, point platoon assigned to the manned company and battalion command group elements. In contrast to the full-mission scenario, the DCE's comprise a series of offensive maneuvers that force this point platoon to respond rapidly to a fluid battlefield situation. The DCE manning structure targets the vertical flow of information between all battalion echelons from individual vehicle commanders to the battalion command level. Measures of performance include throughput time and accuracy for battlefield communications transmitted across the entire battalion and the manned platoon's ability to effectively execute impromptu mission changes. As with the full-mission scenario, the remainder of the battalion's combat vehicle assets are simulated by BLUFOR units for execution of the DCEs.

Data Collection Exercises

ARI's emphasis on robust measures of soldier-in-the-loop performance was the primary catalyst for development of the DCEs. Despite the structured nature of the scenarios used in the CVCC company and battalion evaluations, extended operations during a combat mission are subject to free-play arising from factors such as the direction, speed, formation, and attrition of the opposing units. The DCEs, therefore, are a series of mission segments or "snapshots" from an operational scenario designed to standardize battlefield conditions and soldier placements at critical points and times.

Evaluation Issues and Measures

For more structured and explicit research issues, the DCEs are based on the tactical Battlefield Operating Systems (BOS) of Maneuver, Fire Support, Command and Control, and Intelligence provided in the Blueprint of the Battlefield (Department of the Army, 1991). The research issues address each of these systems. For example, what is the CVCC system's impact on the Command and Control BOS? The research hypotheses are stated at the generic function or task level for each BOS and propose that CVCC units perform significantly better than conventionally equipped units.

Measures for the DCEs systematically cover the BOS functional areas indicated and establish unit linkage between echelons up to battalion. Whenever appropriate, measures at each echelon are cumulated to develop overall battalion performance measures. For example, for the Command and Control BOS function Receive and Transmit Enemy Information, mean response times for reporting enemy target or indirect fire are initially obtained from tank commanders in the point platoon. Mean relay times at each echelon track the flow of this information through the battalion, and finally cumulative times for communicating enemy information are calculated for the entire battalion.

Operational definitions are established for each of the 69 measures identified for the DCEs. The CCTB's Data Collection and Analysis system automatically captures many of these measures, such as unit maneuver and commander's usage of the instrumented CVCC command and control display for tactical communications. Procedures and instruments required for manual data collection, particularly for voice communications by conventionally equipped units, are developed and include report transcription and behavioral observation.

Performance Requirements

To obtain repeated measures for the CVCC research issues identified, the DCEs require the manned participants to perform a series of tactical exercises (Table 1). Each exercise is based on a fragmentary order (FRAGO) that forces an impromptu change in the battalion's original Movement to Contact mission. Typically, each exercise requires the manned participants to: maneuver and navigate while processing direct and indirect fire targets; receive and transmit mission, friendly, and enemy information; and collect enemy information.

The manning structure for the DCEs results in a fully manned platoon (4 simulators) assigned to a flank unit such as B Company, the B Company commander and his executive officer (2), and the battalion commander and his operations officer (2). The two semiautomated platoons of B Company are tethered to the manned platoon. B Company is positioned on the battalion's flank and the manned platoon is designated the point element for this company. Manning structure combined with teleportation to

Table 1

Overview of Data Collection Exercises

Description	Purpose	Cue Event	Direction	FRAGO
Cross Reinforce	Training	Brigade Order	Top Down	Overlay
Bypass Enemy	Training	Intelligence	Top Down	Overlay
Assault Enemy	Evaluation	Intelligence	Top Down	Overlay
Enemy Flank Attack	Evaluation	Enemy Fire	Bottom Up	Oral
Abort an Attack	Evaluation	Friendly Loss	Bottom Up	Oral
Pressured Withdraw	Evaluation	Intelligence	Top Down	Overlay

prespecified locations at the start of each exercise help ensure that participants are located at the critical point and time to require their execution of each FRAGO.

Each exercise is triggered by a cue event, see Table 1, that requires a modification to the original mission. Cue source is either a "top down" event alerting the battalion to a change in the battlefield situation such as a brigade intelligence report updating the enemy's disposition, or a "bottom up" event detected and reported by B Company such as enemy contact. Cue events are designed to force the issue of a prespecified FRAGO directed primarily at B Company. Time permitting in the operational context, graphic overlays detailing the FRAGO are transmitted to CVCC participants on their command and control display. Oral FRAGOS are issued over voice radio.

Standard Operating Procedure (SOP)

The battalion evaluation schedule requires one week of support from each group of test participants, either CVCC or conventional. The first three days are spent in training and provide participants a structured sequence of training events that address: each crew's required use of the simulator and its equipment including CVCC components; CCTB technologies including teleportation, tethering, and automated reporting; and, unit-level rehearsal of simulated operations in the execution of training scenarios. During the fourth day each test unit conducts the full-mission defensive scenario used for horizontal slice evaluation, and on the final day the unit completes the DCEs for vertical slice evaluation.

The conduct of the DCEs is directed by predefined operating procedures designed to standardize key aspects of participant and research personnel performance. The battalion operations order for the original mission includes a unit SOP and is briefed and distributed in the same manner to CVCC and conventional units. Prior to reception of this OPORD, participants are briefed on the intent, nature, and schedule of the DCEs.

This DCE briefing includes oral and slide coverage of the required manning structure, radio net configurations, movement speed and dispersion standards, and company formations. The manned platoon's role as the point element, for example, is emphasized by graphic depictions of company column, wedge, line and other formations. Task standards such as displacement range and the percent of casualties prompting an aborted attack, based on battalion-level Mission Training Plans (MTPs), are reinforced. Participants are instructed that the battlemaster's directive to "Cease Fire, Freeze" indicates they have completed the exercise or expended the time allotted, typically 30 minutes.

For implementation of the DCEs in ARI's ongoing battalion evaluation, two of the six exercises developed are used as training exercises (Table 1). Training objectives for these exercises are addressed in the DCE brief and the unit's performance is reviewed as each exercise is concluded. The objectives stress the maneuver and communication performance requirements on which the DCEs are based.

Prior to each exercise, all manned and semiautomated simulators are teleported to prespecified battlefield locations. The exercise begins when the unit issues a "REDCON 1" indicating ready. As the exercise develops, the battlemaster and OPFOR operator ensure occurrence of the scripted cue event. After FRAGO issue and dissemination, BLUFOR operators maneuver their forces at the direction of the B Co and battalion commander as the unit attempts to complete the FRAGO in the time allotted.

Conclusion

In summary, the DCEs complement the full-mission scenario to efficiently collect soldier-in-the loop, command and control performance data across all battalion echelons. The DCEs reliance on tactical level BOS functions and tasks is expected to provide more meaningful results and a basis of comparison for future evaluations. In conclusion, the DCEs exemplify how others might use simulation-based technologies such as semiautomated forces, tethering, teleportation, and automated battlefield reporting to meet their test and evaluation requirements.

References

Alluisi, E. A. (1991). The development of technology for collective training: SIMNET, a case history. Human Factors, 33(3), 343-362.

Department of the Army (1991). TRADOC Pamphlet 11-9: Blueprint of the battlefield. Fort Monroe, VA: U.S. Army Training and Doctrine Command.

Foss, J. W. (1991, February). Advent of the nonlinear battlefield: AirLand battle-future. Army, pp. 20-37.

Giboney, T. B. (1991, November). Commander's control from information chaos. Military Review, pp. 34-38.

Leibrecht, B. C., Kerins, J. W., & Ainslie, F. M., Sawyer, A.R., Childs, J.M., & Doherty, J. J. (1992). Combat vehicle command and control systems: I. Simulation-based company level evaluation (ARI Technical Report 950). Alexandria, VA: U. S. Army Research Institute for the Behavioral and Social Sciences, Fort Knox Field Unit.

U. S. Army Research Institute
5001 Eisenhower Avenue
ATTN: PERI-RP
Alexandria, VA 22333-5600
(703) 274-5584, DSN 284-5584

THE LONGITUDINAL RESEARCH ON OFFICER CAREERS (LROC): IMPACT OF CHANGE ON ATTITUDES AND CAREERS

Beverly C. Harris

Presented at the 34th Annual Military Testing Association Conference,
28 October 1992, San Diego, California

THE LONGITUDINAL RESEARCH ON OFFICER CAREERS (LROC): IMPACT OF CHANGE ON ATTITUDES AND CAREERS[1]

Beverly C. Harris

U.S. Army Research Institute for the
Behavioral and Social Sciences

Introduction

The last 4 years have witnessed unprecedented world change and turmoil. The cold war ended, Russia and Germany, in particular, have experienced structural, political, and economic instability, and military forces from all over the world came together to stop Iraq's invasion of Kuwait in Operation Desert Shield/Desert Storm (ODS). The United States, during this same period of time, experienced significant economic problems resulting in budget cuts and a call to decrease the size of the military and the Department of Defense. Talk of force reductions were put on hold during ODS; however, at the end of the war, talk of downsizing turned into actions.

The U.S. Army Research Institute has been engaged in a program of research over this same period of time entitled Longitudinal Research on Officer Careers (LROC). The main part of this research has been a survey conducted each year beginning in the fall of 1988. Because of the coincidence of the survey with these significant events, we have the opportunity to track attitudes related to a number of career and Army issues during a period of major change. In addition, for the 1990 and 1991/92 surveys, a specific set of questions was added to address the perceptions of officers on the impact of downsizing on them, their job/career, and on the Army. This paper provides a description of the changes in attitudes and perceptions of the longitudinal group of officers who completed all 4 years of the survey. This group could be considered "survivors" of the first wave of force reductions. Their views on the force-reduction process can inform policy makers and provide important information for "course corrections" during the remaining years of the downsizing.

<u>The LROC Survey</u>

This research takes a longitudinal, life course approach to understanding the career experiences, attitudes, and career decisions of company grade officers from the time they are commissioned. To date the survey has provided annual data: (1) to understand the values, attitudes, family situations, and career experiences of the current generation of Army officers; (2) to test models of the work, career, family and

---

[1] The views expressed in this paper are those of the author and do not necessarily reflect the views of the U.S. Army Research Institute or the Department of the Army.

personal factors that influence career decisions; and (3) to investigate the longitudinal effects of policy change and events on attitudes, perceptions, and career intentions.

Method

Surveys were mailed each year to a stratified random sample of company grade officers [second lieutenants (2LT), first lieutenants (1LT), and captains (CPT)]. Over the 4 years, approximately 10,000 officers have participated in at least one survey. A longitudinal group of respondents (N=928) participated in all 4 years of the survey and are the group analyzed for this paper.

Longitudinal Respondents

Longitudinal respondents were checked against the original sample and the respondents who did not complete all years of the survey. Demographics indicate that the longitudinal respondents are similar to the other respondents and to the original sample. The longitudinal group is made up of 74% male officers and 26% female officers; 84% are white, 9% black, 3% hispanic, and 4% respond "other." Table 1 below shows the change in rank over time.

Table 1

Change in Rank from 1988 to 1991/92 for the LROC Longitudinal Respondents (N=928)

	1988	1989	1990	1991/92
2LT	8%	<1%	--	--
1LT	18%	18%	11%	<1%
CPT	73%	81%	88%	92%
MAJ	--	--	<1%	8%

Results

Perceived Career Prospects

Results indicate that a larger number of officers are concerned about the changes in Army manpower needs and Congressional budget cuts and the implications on their Army career now than in 1988. The percent of officers concerned over changes in manpower needs increased from 39% in 1988 to 72% in 1991/92. Concern over Congressional budget cuts increased from 51% to 73%. Although there were some gender differences in the percentage indicating concern in these two areas from year to year, the overall trend was the same.

In addition, the percent of officers who agreed/strongly agreed they were confident they would be promoted as high as their ability and interest warranted

dropped from 67% in 1988 to 45% in 1991/92. Those who agreed they would get the kinds of assignments they needed to be competitive for promotions also declined from 56% in 1988 to 41% in 1991/92. Overall, approximately 24% of the male officers and 42% of the female officers indicated that the opportunities for command in their branch were limited. Although these overall percentages indicating limited opportunities for command have remained fairly constant for the 4 years, there is considerable variability for both male and female officers by branch ranging as high as 85% in some branches.

Satisfaction With Career Prospects

Overall job satisfaction has remained fairly constant over the 4 years with approximately 78% indicating they are satisfied/very satisfied with their job. Also, about 90% consistently indicate they are proud to tell people they are in the Army. These percentages are similar for both male and female officers. In contrast, satisfaction with their current career prospects has dropped significantly, particularly for male officers. Although females were less likely than males to say they were satisfied in 1988, by 1991/92 there was little difference between them. Table 2 below provides the percentages for each year of the survey.

Table 2

Officers Indicating They are Satisfied/Very Satisfied With Their Current Career Prospects From 1988 to 1991/92 by Gender

	Male Officers	Female Officers
1988	70%	61%
1989	57%	54%
1990	53%	55%
1991/92	50%	51%

Although we can infer that the changes in attitudes reported above may be related to world changes and downsizing, in order to relate attitude change to downsizing directly questions need to specifically address the relationship. In 1990, a set of questions was added to the survey to specifically capture the perceptions of officers regarding the impact of downsizing on them, their job/careers, and on the Army.

Perceived Impact of Downsizing on Job and Career

Over the last two years, 64% of the male officers and 72% of the female officers have indicated that it is likely that they will work more hours as a result of downsizing. It is important to note that these officers already report working an average of 57 hours a week. In 1990, 34% of the male officers and 29% of the female officers thought that it was likely they would suffer because of downsizing; these percentages increased in 1991/92 to 41% for males and 35% for females. In addition, the percent

who thought it was likely that their family would suffer also increased in the two years (for male officers it went from 32% to 35%; for female officers from 20% to 27%).

When asked how likely it was that they would be promoted on or ahead of schedule during a time of force reductions, fewer than 40% indicated that it was likely/very likely. Fewer male officers than female officers thought it was likely they would be promoted on or ahead of schedule in both years (about 31% of male officers for both years; 38% of female officers). In both years, approximately 18% of the males and 19% of the females thought it was likely/very likely they would be involuntarily released from the Army.

In 1990, 29% of the male officers and 35% of the female officers indicated that the probable reductions in the size of the Army made them less/much less interested in staying in the Army now compared to a year ago; in 1991/92 these percentages increased to 39% for males and 42% for females.

Perceived Impact of Downsizing on the Army

Figure 1 displays graphs for five questions on the survey that addressed the potential impact of downsizing on the structure and function of the Army. Results indicate a decreasing trend in the percentage of officers who believe that the "Best" soldiers will stay in the Army; and an increasing trend in the percentage who believe that morale and readiness will suffer as a result of force reductions.

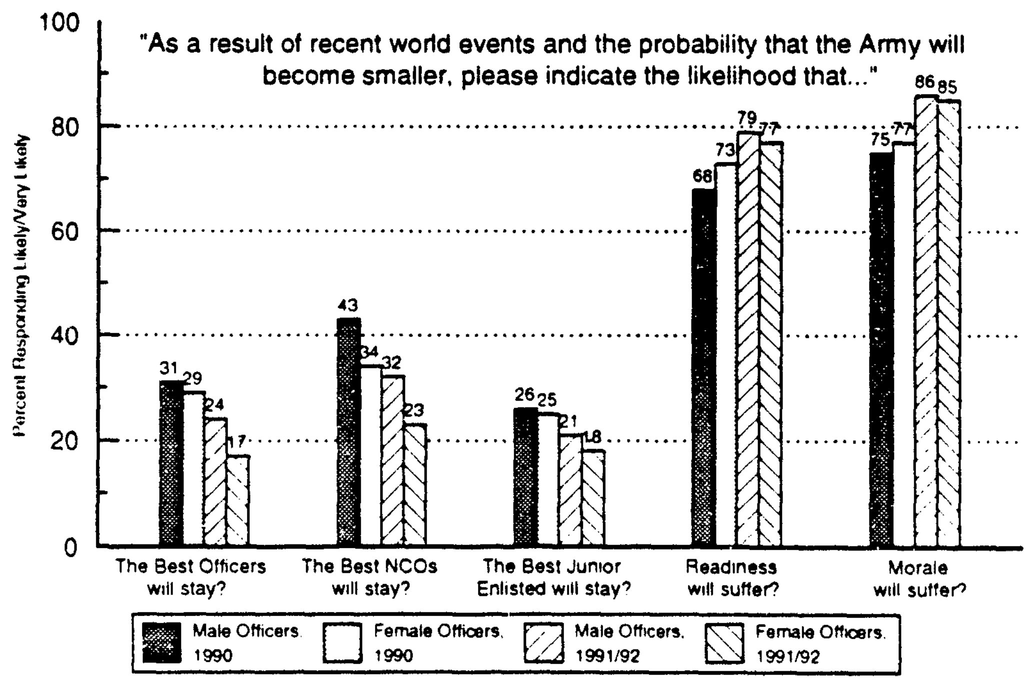


Figure 1. Company grade officers' opinions of the impact of downsizing on the Army

Conclusions and Implications

The findings reported above indicate that the longitudinal respondents who have "survived" the first wave of force reductions are very concerned about the impact of downsizing on them, their job/career, and on the Army. This trend toward more negative attitudes is counter to past research which generally finds a trend toward more positive attitudes with more time in service and increased rank.

Many of the attitudes expressed seem fairly realistic, such as the expectation they will work longer hours in a time when there are fewer people to accomplish the same workload. Also, the perception that promotions will be slower or that certain jobs may not be available may also be realistic in a time of budget cuts. However, it is difficult to judge how realistic the perception is that the "best" are leaving the Army and that morale and readiness will suffer.

Whether or not the perceptions expressed by these "survivors" accurately reflect reality, they can influence behavior, motivation, and morale. In turn, these attitudes may influence career intentions, performance, and future readiness. The Army seems to be providing information and showing concern for the soldiers who are leaving or are being involuntarily separated. However, from the attitudes and concerns expressed in the survey, the Army may need to put more effort into encouraging "high-quality, high performers" to stay, communicating concern for the short-term negative effects of downsizing on the "survivors," and providing more information on "the final product"--the structure and function of the Army after downsizing.

The Validity of Two Methods of Testing Flexibility

Ronna F. Dillon and Rodney J. Greer
Southern Illinois University

Cognitive flexibility becomes increasingly important as the demands of society diversify. To function effectively in a society that places increasingly complex and novel demands on its members, one must possess a superordinate set of abilities involving cognitive flexibility. Dillon (1992) conceptualizes cognitive flexibility as a three-component superability. One component of flexibility identified by Dillon (1988) involves the capacity to generate multiple solution protocols during solution of complex reasoning items that can be solved in more than one way. This dimension of flexibility, called "flexible combination," is the subject of the experiment reported in this paper and is seen in the broader context of the three domain model of flexibility conceptualized by Dillon.

A Componential Subtheory of Cognitive Flexibility

The three-component model of cognitive flexibility can be viewed in the context of Sternberg's (1985) Novelty Subtheory of his Triarchic Theory of Intelligence. The first component of this subtheory involves selective encoding, which in the present context, we term "flexible encoding." Flexible encoding is oparationalized as the ability to encode stimulus attributes flexibly (i.e., in more than one way), ascertaining which stimulus attributes are relevant, or encoding flexibly the meaning, nature, or function of a stimulus attribute.

The second component, **flexible combination,** is operationalized as the ability to assemble item elements in more than one distinct manner when given complex reasoning items that can be solved using more than one tactic. Such assembly requires the ability to generate multiple rules of inference for solution of items where a single correct answer can be induced using multiple tactics.

The third component, **flexible comparison**, is operationalized as the ability to compare flexibly item element configurations to ascertain the most effective assembly of components. Flexible comparison is assessed by comparing information-processing efficiency for items that are similar in inference to a set of items just presented versus an item that follows this set and that is different in inference from the set of items that just preceded it. Such comparison implies relating newly acquired information to previously acquired information.

In previous research, Dillon (1988) reported that the magnitude of the variance in academic, vocational, or professional training success accounted for by flexible combination was inversely related to the amount of structure inherent in the particular training domain. Technical career training, undergraduate psychology and educational psychology, and medical school samples were compared. The variance in academic, technical or training success accounted for by flexible combination was least important for performance in technical career training and most important in third-year medical school, where 53% of

the variance in final exam performance was accounted for by flexible combination. The results are as expected, considering the strong emphasis on low structure problem-solving in the third and fourth years of medical school.

Also in previous research, Dillon and Brannan (1991) provided data to demonstrate the construct validity of flexible combination, relating this intellectual superability to different aspects of intellectual style. This earlier report also contains data relevant to the criterion-related validity of flexible combination, demonstrating that flexible combination predicts cumulative college grade-point average, previous high school rank, and previous ACT scores.

The experiment reported in this paper is designed to test the premises that (a) subjects demonstrate greater flexible combination ability when flexibility is assessed using an elaborated testing method than under a nonelaborated testing procedure, and (b) flexible combination accounts for significant proportions of variance in academic achievement when tested under either a nonelaborated condition or an elaborated procedure.

Subjects

Two hundred eighty-seven college undergraduates served as subjects in this experiment. Males and females were approximately equally represented.

Instruments

As the primary measure of cognitive flexibility, twelve complex figural analogies were taken from the Advanced Progressive Matrices (Raven, 1962). Each item selected met the criterion that it could be solved using at least three tactics. Regardless of tactic, only one response option was correct for a given item. Cumulative grade-point average and ACT Math score also were secured for each subject.

Procedure

Subjects were randomly assigned to one of two conditions: (A_1) standard, and (A_2) elaborated. For the standard flexibility condition, subjects were instructed to generate as many different written protocols for solution of a given test item as possible. Subjects were informed that they would receive a flexibility score from the total number of solution protocols, not by the number of items attempted overall. Under the elaborated condition, possible solution protocols were presented in a 10-minute demonstration prior to the onset of testing. The sample item could be solved using at least five solution protocols.

Results

During the first phase of analysis, the general linear model was used to test the premise that subjects participating in the elaborated method of testing flexibility scored higher on the flexibility measure (defined by the number of strategies generated divided by the number of

items attempted), $F(1, 284) = 11.92$, $p < .001$. During the second phase of analysis, data were analyzed to ascertain the criterion-related validity of each method of testing flexibility. Data indicate that the nonelaborative method of testing flexibility predicts cumulative grade-point average, $F(1, 146) = 6.54$, $p < .05$. Similarly, the nonelaborative approach predicts ACT Math, $F(1, 75) = 6.66$, $p < .05$. Using the elaborative method of testing flexibility, data indicate that flexibility predicts cumulative grade-point average, $F(1, 138) = 4.07$, $p < .05$. Moreover, flexibility tested under the elaborative condition predicts ACT Math, $F(1, 71) = 6.09$, $p < .05$. Means and standard deviations for all measures are presented in Table 1.

Discussion

Cognitive flexibility has been demonstrated to be an important superability in predicting measures of academic success. Data from this experiment indicate that flexibility is enhanced through dynamic test instructions, and that both methods of testing flexibility predict external measures of academic achievement.

Perhaps, subjects tested under the elaborative procedure vary significantly in the extent to which they attend to or benefit from the elaborative techniques modeled by the experimenter. Some subjects are able to benefit from (i.e., become increasingly flexible) as a result of exposure to examples of flexible information processing, while the flexibility level of other subjects is more deeply entrenched. We are suggesting that sensitivity to flexibility instruction is an individual difference dimension; a dimension that should be related to learning (i.e., ability to profit from experience) in general.

An interesting individual difference analysis that would address the above premise would involve ascertaining the extent to which an individual's increases in flexibility under elaborative conditions relative to nonelaborative testing procedures are related to external measures of learning ability. Research directed toward this question is underway.

References

Dillon, R. F. (1988, November). Flexibility in three domains. Paper presented at the Annual Convention of the Psychonomic Society, Boston.

Dillon, R. F. (1992). A componential subtheory of cognitive flexibility. Technical Report 92-1. Carbondale, IL: Cognitive Processes Research Project.

Dillon, R. F., & Brannan, T. S. (1991, November). Toward construct validation of cognitive flexibility. Paper presented at the Annual Convention of the PSychonomic Society, Boston.

Raven, J. C. (1962). Advanced Progressive Matrices, Set I. New York: Psychological Corporation.

Sternberg, R. J. (1985). Beyond IQ: A triarchic theory of human intelligence. New York: Cambridge University Press.

Table 1

Means and Standard Deviations for All Measures

Standard Method of Testing Flexibility*

Variable	**Mean**	**Standard Deviation**
GPA	2.99	.52
ACT	19.66	4.14
Flexibility	.71	.35

*n= =147

Elaborated Method of Testing Flexibility**

Variable	**Mean**	**Standard Deviation**
GPA	2.95	.60
ACT	19.44	3.89
Flexibility	1.27	.68

**n = 140

The Validity of Information-Processing Parameters of Inductive Reasoning

Ronna F. Dillon and Catherine A. Harris
Southern Illinois University

Measurement of intellectual abilities must be accurate and prescriptively useful if pesonnel selection and classification efforts are to be successful. Despite this clear and everpresent need, traditional psychometric methods fail to predict school, training, or job performance adequately (see Dillon, 1989 for a discussion of shortcomings of traditional testing paradigms and new approaches to testing). Sternberg (1989a) notes the advantages that can be realized when information-processing componential approaches are used to derive information about cognitive abilities. One componential approach that has been very productive for increasing criterion-related validity as well as providing information for selection purposes, is the information-processing approach reported by Dillon (1985, 1986, 1989, 1991, 1992a). Using this approach, information-processing components are derived directly from the complex reasoning tasks psychologists seek to understand. As Sternberg (1989b) notes, "Dillon's work helps us understand, quite directly, just what people do when they process information in the performance of intellectual tasks."

Two major classes of paradigms have been used recently to study information-processing componential abilities. Paradigms involve either use of physiological equipment to record ongoing information-processing activity during solution of intact tasks (see Dillon, 1989b for a review of work using eye movement measures of information processing) or some computer-administered means of deriving isolated cognitive processes from administration of the same complex tasks (Dillon & Tirre, 1988). Both methodologies yield data regarding individual differences in information-processing abilities at four levels: (A) **Stages** of information processing, including measures of the number of times components are executed, and the latency for each execution of each component; (B) the **sequential distribution** of processing steps, such as the percentage of components executed in the main stimulus array prior to the first break in ongoing processing to attempt response selection; (C) **strategies or strategy components**, such as image rotation or double-checking; and (D) **adaptations over time**, including measures of the information-processing substrates of learning and flexibility. The learning level of individual differences permits linking construct and criterion-related validity in an interesting way. Not only do information-processing parameters of learning predict external measures of training and academic success (i.e., criterion-related validity), but differences in information-processing efficiency at the beginning versus the end of a trial block reflect interpretable changes in information processing as well. Such learning changes provide evidence of construct validity. The computer-administered, multiple-screen paradigm is employed as a low-cost method of isolating information-processing components, which provides information that is analogous to component information derived from the eye movement paradigm.

Individual differences in the execution of distinct information-processing stages were quantified, as were the order in which processing steps were executed, and changes in information processing over time. This last level of individual differences provides measures of the information-processing substrate of learning. In quantifying learning in this way, differences over trials in the execution of information-processing stage and sequence indices are computed. These learning indices then are used as predictors of the criteria of interest.

The study reported in this paper had two specific purposes. First, construct validity information was presented by demonstrating that interpetable differences in information-processing components were found as a function of learning. Second, evidence of criterion-related validity was presented by validating the information-processing indices against academic and medical preparation performance.

Method

__Subjects__. Two samples were included in this study. Sample 1 was comprised of 11 graduate students enrolled in a 2-year medical preparation program. Sample 2 was comprised of 139 college undergraduates. Examinees ranged in age from 21-35 years old. In addition, 467 beginning Airmen provided the construct validation data.

__Instruments__. Sixty-four items from the Reasoning Battery (Dillon, 1992b) were used as stimuli. The test contains eight each of eight different item types: Figural and verbal classifications and analogies, governed by semantic or nonsemantic inferences. All items of a particular type were presented in a trial block. Academic or medical school performance data also were collected for each examinee.

__Procedure__

Examinees worked at computer workstations. Each item was decomposed into separate information-processing components. For an item, each component was presented on a separate computer screen. Examinees control the amount of time spent processing each screen as well as movement across screens. During analysis, data were extracted with respect to the number of times each stage was executed, the amount of time spent executing each stage, the amount of time spent executing each stage, the sequential ordering of processing stages, and other measures of processing efficiency, learning, and cognitive flexibility.

Results

With respect to construct validation, data were analyzed using the to ascertain the extent to which interpretable differences in information-processing componential abilities exist as a function of learning. Learning is operationalized as increases in the efficiency with which information-processing operations are executed, over a block of eight items of the same type and content. Note, of course, that the execution of certain indices increases with learning, while other components are executed less frequently with learning. The magnitude

and direction of the difference in each index between the first and last items in each block of trials was computed. Three examples of the many effects of learning on information processing are (a) examinees execute a greater percentage of their processing resources during encoding/inference activities, (b) examinees execute a smaller percentage of their processing resources in confirmation activities, and (c) examinees show less redundancy over trials in a given trial block, by executing components a smaller number of times. Results of these and other analyses pertaining to the information-processing substrate of learning are presented in detail elsewhere (Dillon, 1992c).

Regarding criterion-related validity, the information-processing componential abilitiesa are validated against cumulative grade-point average and ACT composite score. Models are tested at the stage level of individual differences and at the sequence level of individual differences, both summed across all items in all trial blocks as well as comparing processing between the first item in each block of eight trials against the last item in each block of trials. These latter comparisons provide data regarding the information-processing substrate of learning. Examples of these analyses are presented below. Analyses are presented in detail elsewhere (Dillon, 1992b).

With respect to stage analyses for the college students, data indicate that 22% of the variance in academic performance is accounted for by a model comprised of the number of times encoding/inference, rule applcation, and confirmation components are executed, $F(3, 85) = 8.01$, $p < .001$. The same model was tested for learning across items within the eight trial blocks. Data indicate that 16% of the variance in academic performance is accounted for by the model, $F(3, 135) = 8.79$, $p < .001$. Testing a stage model on the medical preparation sample, data indicate that 75% of the variance in medical preparation grade-point average is accounted for by a model comprised of the number of times encoding/rule inference and rule application components are executed, $F(2,8) = 11.92$, $p < .01$. As an example of a model that combines stage parameters with sequence level of individual differences, data indicate that 54% of the variance in medical preparation performance is accounted for by a model comprised of the number of times encoding/inference processes preceed rule application processes and the percentage of the total components executed that are encoding/inference processes, $F(2, 8) = 4.66$, $p < .05$. Learning indices of information-processing also were computed for medical preparation students. As an example of a mixed model, containing stage and sequence indices, 81% of the variance in grade-point average for medical preparation students is accounted for by a model tapping the percentage of total components that are confirmation components and the percentage of times encoding/inference processes are executed before rule application processes, $F(2, 8) = 16.91$, $p < .01$.

Discussion

The results of this study indicate that the multiple-screen computer-generated paradigm is a useful way of deriving data about distinct information-processing operations. Moreover, components derived from this approach are powerful predictors, when validated against measures of academic and training success for undergraduate

college students and students pursuing careers in medicine. In addition to the criterion-related validity of this work, the paradigm makes it possible to acquire important data regarding the information-processing strengths and limitations underlying an individual's reasoning task performance.

References

Dillon, R. F. (1985). Predicting academic achievement with models based on eye movement data. Journal of Psychoeducational Assesement, 3, 157-165.

Dillon, R. F. (1986). Information processing and testing. Educational Psychologist, 20(3).

Dillon, R. F. (1989). Information processing approaches to intelligence. In R. J. Sternberg (Ed.), Advances in the psychology of human intelligence (Vol. 5). Hillsdale, NJ: Erlbaum.

Dillon, R. F. (1991). New approaches to aptitude testing. In R. F. Dillon & J. W. Pellegrino (Eds.), Testing: Theoretical and applied perspectives. New York: Praeger.

Dillon, R. F. (1992a). A computerized paradigm for isolating information-processing parameters of intelligence. Learning and Individual Differences. Manuscript submitted for publication.

Dillon, R. F. (1992b). Psychometric credibility of the Reasoning Battery. Unpublished manuscript.

Dillon, R. F. (1992c). Construct validation of information-processing approach to understanding intelligence. Carbondale, IL: Tech. Rep. 92-2), Cognitive Abilities Measurement Program.

Dillon, R. F., & Tirre, W. C. (1988, April). Componential and metacomponential approaches among Airmen. Paper presented at the Annual Convention of the American Educational Research Association, Boston.

Sternberg, R J. (1989a). Componential analysis: A recipe. In D. K. Detterman (Ed.), Current topics in Human intelligence (Vol. 1). Norwood, NJ: Ablex.

Sternberg, R. J. (1989b). Advances in the psychology of human intelligence (Vol. 5). Hillsdale, NJ: Erlbaum.

Further Construct Validation of Cognitive Flexibility

Ronna F. Dillon and Timothy S. Brannan
Southern Illinois University

The important role of flexibility in today's multifaceted technical job environments is clear. Job analyses of technical tasks invariably contain lists of knowledge, skills, abilities, and other attributes that reflect abilities to encode stimulus attributes flexibly, using multiple inferences regarding element transformations to solve items, and maintain information-processing efficiency as the demands of a task change continually.

Dillon's (1992a) three-component model of cognitive flexibility can be viewed in the context of Sternberg's (1985) Novelty Subtheory of his Triarchic Theory of Intelligence. The first component of this subtheory pertains to selective encoding, which in the present context, we term "flexible encoding." Flexible encoding is operationalized as the capacity to encode flexibly the attributes of a stimulus (i.e., encode stimulus attributes in more than one way; formulate more than one definition for a given stimulus or stimulus attribute, provide more than one meaning for a given stimulus, create more than one configuration for a set of stimulus attributes).

The second component of the Novelty Subtheory is "selective combination," which in this context we term "flexible combination." Flexible combination is operationalized as the capacity to assemble (i.e., select and combine) item elements in more than one distinct manner when presented complex reasoning items that can be solved using more than one tactic. As Sternberg notes, such selective or flexible combination involves putting together pieces of information that might originally seem isolated into a unified whole. This unified whole may not resemble its parts.

The third component of the novelty subtheory is selective comparison, which in the present perspective we term "flexible comparison." Flexible comparison is operationalized as the capacity to flexibly compare item element configurations to ascertain the most effective assembly of components when presented more than one potentially appropriate assembly. Such comparison implies relating newly acquired information to previously acquired information. The problem solver must understand how the newly presented information is similar to and different from information presented in the past, and how different examples of newly presented information are distinct from one another. Flexible comparison must be assessed by determining the extent to which the subject maintains information-processing efficiency when items governed by new types of inferences follow sets of inferentially similar items. For the three studies reported in this paper, interest is in the second component of flexibility; i.e., the ability to generate multiple strategies in solution of complex reasoning tasks that can be solved in more than one way.

Dillon's (1988, 1992b) flexibility dimension of flexible combination has been found in earlier work to play a very important role in academic and training success for a wide range of populations. In previous research, Dillon and her colleagues (Dillon, 1988, 1992b; Dillon & Brannan, 1991) reported that flexibility accounts for significant amounts of variance in academic performance for undergraduate college students, graduate students in medical preparation courses, and third-year medical students enrolled in a psychiatry clerkship.

The learning/study style underpinnings of flexible combination were elucidated in other construct validity research (Dillon & Brannan, 1991). Elaboration (i.e., breadth of processing) and personal attribution were found to be related to flexible combination, while depth of processing, as expected, was unrelated to flexibility. The present study goes a step further in including measures of intellectual style type along with learning/study and attribution processes to refine our understanding of the nature of flexible combination.

The present study is designed to provide evidence of construct validity for the flexible combination dimension of cognitive flexibility. Specifically, data will be reported that demonstrate that this dimension of flexibility subsumes aspects of intellectual style type, learning/study processes, and attribution. With regard to the learning/study processing underpinning of cognitive flexibility, we hypothesize that elaborative processing will account for significant variation in flexible combination, whereas depth of processing will not predict flexible combination. Moreover, with regard to the intellectual style type substrate of flexibility, we hypothesize that a style type of Intuition is related to flexibility because both dimensions necessitate the ability to make multiple inferences about relations among phenomena. A final piece of construct validation data comes from the relationship of locus of control to flexible combination. Here we use the second definition of flexible combination; the total number of strategies attempted. The reasoning here is that the more internal one's locus of control, the more likely one will execute a large number of solution protocols. Criterion-related validity data also will be presented.

Method

Sample

The sample was comprised of 130 college undergraduates, ranging in age from 29-35 years old. Males and females were approximately equally represented.

Instruments

Twelve items were taken from the Advanced Progressive Matrices (Raven, 1962). Test items are 3 x 3 figural analogies, each having an 8-item response set. Each analogy item was selected because it could be solved in at least three different ways to produce the single correct answer. Breadth of processing was tapped using the Learning Style Inventory (Schmeck, Ribich, & Ramaniah, 1977), a self-report study

skills instrument. Intellectual style type was assessed using the self-report Form G of the Myers-Briggs Type Indicator (Briggs & Briggs-Myers, 1962). The Indicator is a self-report inventory that measures preferences on the dimensions of extraversion-introversion (E-I), sensing-intuition (S-N), thinking-feeling (T-F), and judgment-perception (J-P). The introversion-extraversion dimension differentiates people who prefer to orient their perception and judgment toward the external world of people and objects, while people who prefer to use introversion are concerned primarily with the inner world of ideas and concepts. The judgment-perception dimension differentiates people who prefer a judging attitude toward the outside world from individuals who prefer a perceiving attitude. The sensing-intuition dimension distinguishes people who prefer to perceive directly through the senses versus a less obvious process of intuition which yields meanings, relationships, and possibilities. The thinking-feeling dimension reflects a preference for mode of judgment, thinking being an impersonal local process or impression and feeling being based on relatively personal and social values. The Indicator concerns differences in approaches to perception and decision-making. The developers of the Indicator contend that different preferences on these dimensions will be associated with different behavioral and cognitive styles, and that individuals are likely to have become more comfortable and practiced at using a particular orientation. Personal attribution was assessed using the Nowicki-Strickland Locus of Control Scale (Nowicki & Strickland, 1972), a 40-item self-report measure that taps the extent to which examinees believe effort and skill versus luck and external factors are responsible for the events in the examinee's life.

<u>Procedure</u>

Examinees completed the Learning Style Inventory, the Nowicki-Strickland Locus of Control Scale, the Myers-Briggs Type Indicator, and the flexibility measure in an untimed format. Order of administration of the instruments was counterbalanced.

Two flexible combination scores were computed. The first score was a measure of the total number of tactics generated overall, divided by the number of items attempted. The second flexible combination score was a measure of the total number of tactics generated on the items attempted correctly. The first measure of flexible combination is believed to be related to thinking and study skills and to intellectual style type, while the second measure of flexible combination is expected to be related to personal attribution.

Results

Table 1 presents means and standard deviations for flexible combination scores and all intellectual style measures. Data were analyzed using the General Linear Model. A series of construct validation analyses were conducted. With respect to study/learning processes, data indicate that elaborative processing contributes to flexibility (defined as the number of strategies executed divided by the number of items attempted correctly), $F(1, 128) = 6.35$, $p < .05$, whereas, as hypothesized, deep processing does not predict flexibility.

With respect to intellectual style type, data indicate that the personal style type of Intuition predicts flexibility, $F(1, 128) = 8.47$, $p < .01$. Regarding personal attribution, locus of control predicts flexibility, $F(1, 128) = 10.26$, $p < .01$.

Discussion

Construct validation efforts reflect the intellectual style underpinnings of this component of cognitive flexibility; the component that taps flexible combination. Convergent and discriminant validity of learning/study processes provide one source of evidence. As hypothesized, elaborative processing relates to flexibility, while deep processing does not. A second source of evidence comes from the intellectual style type substrate of cognitive flexibility. As hypothesized, a style type of Intuition is related strongly to cognitive flexibility. Both psychological phenomena require abilities to encode or assemble a given set of stimuli in multiple ways, infer multiple rules governing transformations across stimulus elements, and to maintain information-processing efficiency when items governed by new types of inferences follow sets of items governed by a different type of inference. A third source of construct validity data comes from the demonstration that the total number of strategies attempted increases with the degree of personal attribution. The attribution finding is expected because an individual is likely to generate multiple solution protocols for solution of each item to the extent that this individual acknowledges responsibility for the outcome of task performance.

The importance of flexible information processing is well-documented both in job analyses and construct modeling of many military and civilian jobs where large amounts of material must be mastered in concentrated periods of time and wherein trouble-shooting activities are integral parts of job performance. The demonstration of construct validity as well as evidence of concurrent, postdictive and predictive sources of criterion-related validity, in this study and related work, reflect the importance of considering this superability in selection and classification activities.

An additional application of this work centers on the trainability of flexibility. The extent to which individuals profit from instruction (i.e., demonstrate learning) in general could be expected to predict the amount of improvement in flexibility that occurs with flexibility training. The empirical question described here is under investigation.

References

Briggs, K. C., & Myers-Briggs-Myers, I. Myers-Briggs Type Indicator. Atlanta, GA: Educational Testing Service.

Dillon, R. F. (1988, November). Flexibility in three domains. Paper presented at the Annual Convention of the Psychonomic Society, Boston.

Dillon, R. F. (1992a). A componential subtheory of cognitive flexibility. Manuscript in preparation.

Dillon, R. F. (1992b). Construct validity of flexible combination. Manuscript in preparation.

Dillon, R. F., & Brannan, T. S. (1991, October). Toward construct validation of cognitive flexibility. Paper presented at the Annual Convention of the Midwestern Educational Research Association.

Raven, J. C. (1962). Advanced Progressive Matrices, Set I. New York: Psychological Corporation.

Schmeck, R. R., Ribich, & Ramaniah. (1977). The Inventory of Learning Processes. Applied Psychological Measurement.

Sternberg, R. J. (1985). Beyond IQ: A triarchic theory of human intelligence. New York: Cambridge University Press.

Table 1

Means and Standard Deviations for All Measures

Variable	Mean	Standard Deviation
Deep Processing	9.20	1.89
Elaborative Processing	7.82	1.79
Locus of Control	27.69	5.52
M-B Intuitive	.41	.49
M-B N	.48	.50
M-B Thinking	.35	.48
M-B P	.48	.50
Number of Correct Attempts	5.86	2.59
Sum of Strategies	6.89	3.83
RFlex	1.11	.40
Incorrect Attempts	5.68	2.94
Total Items Attempted	11.54	1.21
Flexible Combination B	.62	.40
Percent Correct Attempts	.52	.24

n = 130

Emotional Control, Inductive Reasoning, and Academic Performance

Roger Webb and Ronna F. Dillon
Southern Illinois University

Researchers have documented their interest in the relationship of emotions and behavior since 1890 (Carlson & Hatfield, 1992; James, 1890). A unifying theme in recent research is that emotions serve as powerful organizing forces in human development (Berk, 1991). **Emotion** is "an inferred complex sequence of reactions to a stimulus and includes cognitive evaluations, autonomic and neural arousal, impulses to action, and behavior designed to have an effect upon the stimulus that initiated the complex sequence (Plutchik, 1984, p. 217)."

Studies have centered on the etiologies of emotion (Plutchik, 1980; Schacter & Singer, 1962) and on its expression (Ekman, 1980; Izard, 1971). The work reported in this paper departs from these traditional emphases and focuses on the situationally appropriate **control** of emotion. **Emotional control** is defined as the tendency to inhibit the expression of emotional responses (Roger & Nesshoever, 1987). Moreover, the study reported herein furthers attempts at providing more comprehensive models of behavior, in an academic context, by quantifying the contribution of emotional control to academic performance **above and beyond** the contribution made inductive reasoning.

In previous research in emotional control, the construct was found to predict physiological reactivity and recovery from stress (Roger & Najarian, 1989). The concept is similar to that of the regulation of emotion found in studies of child development, where the ability to control one's emotions has been seen as a hallmark of personal and social development (Campos, Campos, & Barrett, 1989; Kopp, 1989; McCoy & Masters, 1990). The construct of aggression control has been linked to violent behavior among prison inmates (Roger & Nesshoever, 1987).

Method

Subjects

One hundred sixty-six college undergraduates served as examinees. Males and females were approximately equally represented.

Instruments

The Emotional Control Questionnaire 2 (ECQ 2; Roger & Najarian, 1989) is a 56-item, self-report measure, containing four subscales. The subscales tap rehearsal, emotional inhibition, aggression control, and benign control (i.e., control of impulsivity). The ECQ2 is designed to assess the connection between the inhibition of emotional responses and predisposition to stress-related illnesses. The Reasoning Battery (Dillon, 1992) contains nine types of test items varying along item content (i.e., verbal and figural), item type (i.e., analogies, classifications, and syllogisms), and meaning and nonsemantic relations dimensions. The analogy and classification subtests were used in this study.

Procedure

Examinees completed the ECQ2 and the Inductive Reasoning Test (Dillon & Radtke, 1985) in an untimed format, under a computer-generated administration procedure. Academic achievement data also were collected.

Results

Inductive reasoning was validated against cumulative grade-point average, as was emotional control. In addition, the relationship of emotional control to academic achievement was computed, controlling for the contribution made by inductive reasoning ability. In this way, it is possible to demonstrate that this noncognitive construct contributes to academic performance apart from the contribution made by reasoning or intelligence. This finding allows us to counter the statement that people with a high degree of emotional control perform well in school, not because of their control but because they are intelligent and this higher intelligence makes possible greater cognitive control of inappropriate emotional expression. Therefore, we can contend that the noncognitive aspect of emotional control is responsible for the relationship of the construct to measures of academic performance.

Data were analyzed to ascertain the criterion-related validity of inductive reasoning and of emotional control and the significance of the relationship of emotional control to academic success above and beyond the contribution made by general reasoning ability. Means and standard deviations for all measures are presented in Table 1. With respect to the first premise, data indicate that inductive reasoning predicts cumulative grade-point average, $F(1, 164) = 9.01$, $p < .01$, as does emotional control, $F(1, 164) = 11.69$, $p < 001$. Regarding the second premise, data indicate that emotional control contributes to the prediction of academic success, controlling for the significant prediction made by inductive reasoning, $F(1, 164) = 10.35$, $p < .001$.

Discussion

The work reported in this study demonstrates that noncognitive dimensions, such as emotional control, can play important roles in prediction of academic achievement above and beyond the contribution made by cognitive measures such as inductive reasoning. Future work is directed toward the trainability of emotional control.

References

Berk, L. E. (1991). Child development (2nd ed). Boston: Allyn and Bacon.

Campos, J. J., Campos, R. G., & Barrett, K. C. (1989). E,emergent themes in the study of emotional development and emotional regulation. Developmental Psychology, 25(3), 394-402.

Carlson, J. G., & Hatfield, H. (1992). Psychology of emotion. Fort Worth: Harcourt Brace Jovanovich.

Dillon, R. F. (1992). Psychometric credibility of the Reasoning Battery. Manuscript submitted for publication.

Ekman, P. (Ed). (1980). Emotion in the human face. London: Cambridge University Press.

Izard, C. E. (1971). The face of emotion. New York: Appleton-Century-Crofts.

James, W. (1890, 1904). The principles of psychology. New York: Henry Holt.

Kopp, C. B. (1989). Regulation of distress and negative emotions: A developmental view. Developmental Psychology, 25(3), 343-354.

McCoy, C. L., & Masters, J. C. (1990). Children's strategies for the control of emotion in themselves and others. In B. S. Moore & A. M. Isen (Es.), Affect and social behavior. Cambridge: Cambridge University Press.

Plutchik, R. (1984). Emotions: A general psychoevolutionary theory. In K. R. Scherer & P. Ekman (Eds.), Approaches to emotion. Hillsdale, NJ: Erlbaum.

Plutchik, R. (1980). Emotion: A Psychoevolutionary synthesis. New York: Harper & Row.

Roger, D., & Nesshoever, W. (1987). The construction and preliminary validation of a scale for measuring emotional control. Personality and Individual Differences, 8(4), 527-534.

Roger, D., & Najarian, B. (1989). The construction and validation of a new scale for measuring emotional control. Personality and Individual Differences, 10(8), 845-853.

Schacter, S., & Singer, J. (1962). Cognitive, social, and physiological determinants of emotional state. Psychological Review, 69, 379-399.

Table 1

Means and Standard Deviations for Aggression Control, Inductive Reasoning, and Academic Achievement

	Mean	SD	Maximum
GPA	2.80	.56	4.0
Inductive Reasoning	9.16	1.94	14.0
Emotional Control	8.66	2.98	14.0

A Template for Evaluating Studies Presenting Potential Predictors of Effectiveness

William F. Kieckhaefer
RGI, Incorporated

Some of us spend most or all of our time developing new measures. Our intent for those measures might range anywhere from theory development and construct validation to predicting job success. Others of us spend the majority of our time concerned with identifying the best combinations of predictors of job effectiveness. In this presentation, I refer to those of us in this latter group as practitioners. When we are in this practitioner role, we concern ourselves with the appropriate use of developed tests.

In this symposium, the presenters may not have intended their papers and presentations as making contributions for practioners. However, since this is the Military Testing Association, I am reviewing them from that perspective. This is not to suggest that the objectives of theory building are diametrically opposed to the objectives of practice. Rather, I suggest that the purposes of both are met only as we keep a broad perspective.

In reviewing this symposium, I considered four primary criteria: demonstrated theoretical foundation, evidence of observed effect size, evidence regarding subgroup differences, and evidence regarding the effects of faking or training.

Demonstrated Theoretical Foundation

Campbell (1990) and Fleishman, Quaintance, and Broedling (1986) present differing but converging points of view on the usefulness of the theoretical foundations for our research. Few of us contend their arguments, and few studies in our discipline fail to meet this criterion in the general sense. Yet Campbell (1990) still writes of our "underspecified steps in the deduction of the prediction from the theory" (p. 47) and calls for greater specificity in our hypotheses. Certainly, our field drives this point into its graduate students and junior professionals early in their careers.

Practitioners require this information for the same reasons basic researchers and theoreticians need it (Fleishman, Quaintance, and Broedling, 1986): conducting literature reviews, establishing better bases for conducting and reporting research studies to facilitate their commparison, standardizing study conditions, generalizing research findings, exposing gaps in knowledge, and assisting in theory development. The presenters here today did well both in basing their approaches on available theory and constructs and in explicating their predictions based on theory.

Evidence of Observed Effect Size

In presenting his capital budgeting approach to utility analysis, Cascio (1989) demonstrates how practitioners can use effect sizes from meta analyses to facilitate making decisions about investments in human resource programs. While his chapter specifically addressed investments in training programs, the decision processes easily apply to investments in personnel selection programs.

As time goes on, we find ourselve in our practioner roles making more justifications to our sponsors not only in theoretical terms but also in cost-benefit or utility terms. We can more effectively make these arguments at the outset to the extent we have information available to us regarding expected effect sizes associated with using one predictor versus another.

Hunter and Schmidt (1990, pages 271-274) present several formulas demonstrating the equivalence of several types of information from which a reader can compute effect sizes. Essentially, if researchers do not include effect sizes, they should provide one of the following combinations of statistics to enable a reader to compute one:

ANOVA's Means and pooled, within-group SDs, or
Each group's M, SD, & N, or
r, or
t and N.

Here, too, the presenters did well in that they presented each group's mean, standard deviation, and sample size.

Evidence Regarding Subgroup Differences

Hunter and Hunter (1983) cite meta-analyses and validity generalization work to support their claim that ability tests which are valid for one subgroup are also valid for the others. Arvey and Faley (1988) support this point of view. Unfortunately, the same body of research they cite suggests considerable adverse impact on subgroups due to real subgroup differences in mean scores on both predictors and performance criteria.

To the contrary, Reilly & Warech's (1988) review shows that typical behavior measures (like biodata) tend to result in relatively low levels of adverse impact. Similarly, Dillon (1981) argues that the nature of the eye-movement measurement stimuli coupled with the use of computer-driven ongoing psychophysiological recording greatly reduces any cultural bias that may exist using the testing of intelligent performance.

Practitioners are seeking valid predictors which show little or no adverse impact. It follows that applied predictor development work must focus research efforts on delineating the

types of predictors and the occupational settings most likely to yield reduced adverse impact. In all fairness, the presenters in today's symposium had little opportunity to investigate subgroup differences given the small sample sizes in their studies.

Evidence Regarding the Effects of Faking or Training

There does exist research to support guarding against coaching on cognitive tests. For example, Boldt, Centra, and Courtney (1986) demonstrate that initial scores on the Scholastic Aptitude Test (SAT) predict college performance better than the highest SAT scores. Navy researchers found similar results using the SAT to predict performance in the Naval Academy.

With typical behavior measures like biodata, faking is the issue rather than trainability. For these types of predictors there does exist a body of research (eg., Abrahams, Neuman, and Githens, 1971; Cascio, 1975; and Trent, 1987) describing the low probability of faking when tests include appropriate types of items.

The question here for those of us developing new predictors or refining theoretical constructs is this: "When examinees receive coaching or training on these constructs, does the increased test performance reflect real increments in ability leading us to correctly predict improved job performance?" Several of the papers presented in this symposium discuss future research planned to assess the trainability of the cognitive abilities investigated.

REFERENCES

Abrahams, N.M., Neuman, I., & Githens, W.H. (1971). Faking vocational interests: Simulated versus rel life motivation. Personnel Psychology, 24, 5-12.

Arvey, R.D., & Faley, R.H. (1988). Fairness in selecting employees (2nd ed.). New York: Addison-Wesley Publishing Company.

Boldt, R.F, Centra, J.A., & Courtney, R.G. (1986). The validity of various methods of treating multiple SAT scores (College Board Rep. No. 86-4). New York: College Entrance Examination Board.

Cascio, W.F. (1989). Using utililty analysis to assess trainilng outcomes. In I. Goldstein (Ed.), Training and development in work organizations: Frontiers of industrial and organizational psychology. San Francisco: Jossey-Bass.

Campbell, J.P. (1990). The role of theory in industrial and organizational psychology, in M.D. Dunnette & L.M. Hough (Eds.) Handbook of industrial & organizational psychology (2nd Ed, Vol. 1). Palo Alto, CA: Consulting Psychologists Press, Inc.

Cascio, W.F. (1975). Accuracy of verifiable biographical information blank responses. Journal of Applied Psychology, 60, 767-769.

Dillon, R.F. (1981). Individual difference in eye fixations within and between stages of inductive reasoning. (Tech. Rep. No. 2) Carbondale, IL: Southern Illinois University.

Fleishman, E.A., Quaintance, M.K., & Broedling, L.A. (1984). Taxonomies of human performance: The description of human tasks. San Diego: Academic Press, Inc.

Hunter, J.E., & Hunter R.F (1983). The validity and utililty of alternative predictors of job performance. (OPRD Report 83-4). Washington, D.C.: U.S. Office of Personnel Management.

Hunter, J.E., & Schmidt, F.L. (1990). Methods of meta-analysis: Correcting error and bias in research findings. Newbury Park, CA: Sage Publications.

Reilly, R.R., & Warech, M.A. (1988). The validity and fairness of alternative predictors of occupational performance. Paper submitted to the National Commission on Testing and Public Policy, Washington, D.C.

Trent, T. (1987). Armed forces adaptability screening: The problem of item response distortion. In L.J. Stricker (Chair), Problems of biodata distortion in personnel selection systems. Symposium conducted at the 95th Annual Convention, American Psychological Association, New York.

ESPIONAGE BY U.S. CITIZENS: 1945-1990

Martin F. Wiskoff
BDM International, Inc.

Suzanne Wood
The Defense Personnel
Security Research Center

Background

In the Stilwell Commission's 1985 report, Keeping the Nation's Secrets (DoD Security Review Commission, 1985), concern was expressed over the increase in reported espionage cases in the 1980s and the lack of research information on espionage and personnel security. PERSEREC initiated a project in 1988 to develop a database on all Americans involved with espionage against the United States since World War II. It was determined at the start that the database should be unclassified in order to allow the widest possible dissemination of information to policy-makers and to others within the government interested in understanding trends in espionage.

A review of the literature found many journalistic and biographical writings about individual American spies, and some compilations of case histories. There was, however, little in the way of systematic attempts to aggregate information across cases and to evaluate patterns and trends among spies. This paper describes construction of a database of American citizens who betrayed their country by providing (or attempting to provide) classified information to foreign powers. It also presents an overall picture of the spies, and information on personal and job characteristics and on the nature of the espionage act itself.

Methodology

Criteria for Including Cases

All American citizens allegedly involved in espionage between 1945 and 1990 for whom unclassified sources of information were available were reviewed. Sources consulted were newspaper and magazine articles, biographies of spies, general descriptive works on espionage, and other researchers' synopses of cases. Over 150 individuals were identified from these sources as potential espionage cases. Upon review of these cases, the following criteria for inclusion in the database were developed: (1) Individuals convicted of espionage or for attempting or intending to commit espionage; (2) Individuals prosecuted for espionage but who committed suicide before the trial or sentencing could be completed; and (3) Individuals for whom clear evidence of espionage (actual or attempted) existed, even though they were not prosecuted. This category included cases involving defections, suicides, deaths, and those administratively processed (e.g., allowed to retire, given immunity, discharged from the military). A total of 117 individuals were identified who met the criteria.

Variable Selection and Coding

Three categories of information were gathered: personal, job and espionage characteristics. Within these categories, variables were selected that might be available from open sources and would

provide a rich array of background data on spies. Included were personal and demographic information, aspects of the spies' job environment, their access to classified information, how they first got involved with espionage, and how their careers as spies evolved and ended. Information on whether they volunteered or were recruited (and by whom) was collected, as were their motivations for committing espionage.

Variables that are subject to change over time were coded according to status at the time when espionage began. For example, marital status was coded according to whether people were married, separated, divorced or single when they started spying.

A total of 56 variables are contained in the database. Some of the variables were included for identifying and documentary purposes only and were not used for analysis. For most of the variables data are available for at least 110 of the 117 spies. There are four variables for which there may be a greater amount of missing data and for which our confidence in their accuracy is lower because of the difficulty of obtaining information from open-source literature. These are immoderate alcohol/illegal drug use, foreign relatives, sexual preference and payment received.

Limitations of the Study

This unclassified study deals only with caught spies whose names surfaced in open-source materials. It is impossible to know how many more spies were caught committing espionage but were not prosecuted for various reasons, or how many have spied in the past and were not caught, or are spying at present and remain uncaught.

Analyses

In conducting the data analyses, frequencies were first calculated for each of the personal, job and espionage characteristics. Next, each variable was explored in relation to the following four major areas of espionage interest:

1. Whether spies differed according to the length of their espionage career. This variable was coded into: (a) the first espionage attempt was intercepted; (b) espionage lasted less than 1 year; (c) espionage lasted 1-4.9 years; and (d) espionage lasted for 5 years or more. People in the latter three categories were termed successful spies.

2. Whether there were differences between uniformed military and civilian spies.

3. Whether spies exhibited different characteristics over time. Time was coded into the decades during which an espionage career began: (a) the half-decade 1945-1949; (b) 1950-1959; (c) 1960-1969; (d) 1970-1979 and (e) 1980 to 1990.

4. How the spies were drawn into espionage: coded into (a) volunteers; (b) those recruited by family or friends, and (c) those recruited by foreign intelligence.

Findings

Table 1 shows that spies were mostly male (108). They were also predominantly white (108) with minorities represented by seven blacks, one Asian-American, and

Table 1

Characteristics of Spies

Characteristics		N cases with data
Gender	Male (108), female (9)	117
Race	White (108), Black (7), other (2)	117
Length of espionage (yrs)	Intercepted (35), < 1 (20), 1-5 (35), > 5 (27)	117
Decade	40s (14), 50s (12), 60s (19), 70s (24), 80s (48)	117
Citizenship	All U.S. (naturalized, 15)	117
Volunteers/recruits	Volunteers (73), family/friends (17), foreign intelligence service (26)	116
Age (yrs)	Median (28.5), range (18 to 69)	116
Education (yrs)	10 (10), 12 (45), 14 (23), 16 (23), 18 (13)	114
Marital status	Married (65), single (39), separated/divorced (11)	115
Sexual preference	Heterosexuals (86), homosexuals (6), unknown (25)	117
Immoderate alcohol/illegal drug use	Alcohol (16), drugs (14), alcohol/drugs (9)	39
Foreign relatives	Yes (41), no (25)	66
Military/Civilian	Military (61), civilian (56)	117
Agencies	Navy (33), Army (22), AF (21), DoD contractors (8), CIA (7), Manhattan Project (6), NSA (5), Marine Corps (4), others (7)	113
Occupation	Commun/intel (35), gen/tech (30), scientific/professional (24), support (18), other (9)	116
Post-employment	Some continued or started after job	117
Security clearance	Top secret (50), confid/secret (30), none (30)	110
Military rank	E1-E3 (13), E4-E6 (30), E7-WO (11), officer (6)	60
Where espionage began	U.S. (76), foreign (35)	111
Foreign country espionage began	W. Germany (14), U.K. (4), Austria (3) Others (1's or 2's)	35
Countries receiving information	USSR (83), E. Germany (7), Poland (4), Hungary (3), Czech (2). Also friendly nations.	117
Payment received	None (47), $50-10K (21), 10K-100K (17), 100K+ (10)	95
Length of sentence (yrs)	0 (20), 1-9 (39), 10-19 (19), 20-40 (22), life (13), death (2)	115
Motivation (Primary)	Money (60), ideology (21), disgruntle/revenge (17), ingratiation (10), coercion (4), thrills (3)	115

one American Indian. Thirty-five individuals were intercepted in their first effort to spy while 82 successfully passed at least some information. The numbers of individuals initiating espionage increased steadily over the decades and then doubled from 24 in the 1970s to 48 in the 1980s.

All the spies were American citizens, in accordance with the criteria for inclusion in the database; however, 15 were naturalized. While most (73) of the spies were volunteers, 17 were recruited by family or friends and 26 by foreign intelligence. Most individuals started their espionage at a young age (median age = 28.5). However, there was an extensive range - from 18 to 69. The largest number of spies (45) had just a high school education although there were 10 who had not completed high school. A considerable number (23) had at least some college, were college graduates (23) or had postgraduate education ((13), including two with doctorates.

When they began espionage, 65 spies were married, 39 single and 11 separated or divorced. Of the 92 for whom sexual preference could be inferred, 86 were heterosexual and six were homosexual. For 25, sexual preference was unknown. Thirty-nine spies were known to have used alcohol immoderately or to have taken illegal drugs. Forty-one had foreign relatives.

There were 61 uniformed military and 56 civilian spies. For those in the military the largest nubmer (30) came from the E4-E6 ranks. There were also 13 younger enlisted personnel (E1-E3s), 11 older E7s or warrant officers, and 6 officers.

The cases were distributed through many agencies, some of which employed both military and civilian workers, with the largest numbers coming from the Navy (33), followed by the Army (22) and the Air Force (21). There were also eight spies employed by DoD contractors, seven with the Central Intelligence Agency, five from the National Security Agency, six associated with the Manhattan Project in the 1940s, four from the Marine Corps, and seven from other agencies.

The occupational areas in which spies were working at the time they began espionage were Communications/ Intelligence (35), General/Technical (30), Scientific/Professional (24) and Functional Support/Administration (18) fields. While most began and ended their espionage while working for the same agency, 26 spies either continued after leaving their place of employment or actually began spying after they had left their primary jobs, sometimes by defecting. The greatest number (50) had top secret clearances, although there were 30 with only confidential/secret clearances and another 30 with no clearances at all; the latter group acquired access to classified materials by various means, such as using go-betweens.

Seventy-six spies began their espionage in the United States compared to 35 abroad. If cases started abroad, the largest number began in West Germany (14). Information was meant for Eastern Bloc countries in 99 of the cases (U.S.S.R., East Germany, Poland, Hungary and Czechoslovakia) and for other hostile nations in another four instances. Friendly or neutral nations were the targets for nine of the spies.

Information is available on the amount of money received for espionage for 95 of the spies. Almost half of these (47) received nothing, because they were discovered before they could be paid or because they acted from nonmercenary motives. Other spies were paid handsomely. For example, 17 received between $10,000 and $100,000, another seven between $100,000 and $1,000,000, and three were paid more than $1,000,000.

The penalties for espionage have ranged from very short sentences, to life and multiple life, to execution. Just over half the spies (59) received either no sentence (20) or less than 10 years (39). There were 13 cases in which life sentences were given, some of which were multiple life.

Money was the most common primary motive (60 cases), followed by ideology (21), disgruntlement/revenge (17), ingratiation (spying in order to please or help someone) (10), coercion (blackmail by foreign intelligence) (4), and thrills/self-importance (3).

Implications

The 1980s, so often called the decade of the spy, produced many young would-be spies who volunteered to commit espionage in exchange for money. The 1980s might well also be dubbed the decade of the intercepted spy, for 28 of the 48 spies (58%) during that 10-year period were intercepted the first time they tried to commit espionage. Even more impressive is the fact that of all the 35 spies intercepted since World War II, 28 were caught in the 1980s.

Despite the fact that many spies were intercepted before they could pass any information, inestimable damage was caused by such groups as the Walkers and the Snowman and the Falcon (Boyce and Lee) and by other individuals.

The data indicate that many of the spies displayed behavior that violated the criteria for being granted and for maintaining clearance and access eligibility. In certain instances these behaviors were directly related to the spies' espionage activities.

The press and the public have probably become indifferent to espionage as it plays out at the national level. Only sensational espionage cases appear to receive much media attention at all. Yet at the working level, it is the awareness of security requirements that in the past has led coworkers to report inappropriate behavior; and this in turn has resulted in the apprehension of several spies.

The data also show that much of the risk of espionage is associated with the type and location of the job a person fills. There are indeed differential risks of espionage associated with overseas assignments, types of occupations and rank. Jobs themselves carry their own vulnerabilities.

Conclusion

With the official end of the Cold War, it might appear to some that espionage is a thing of the past. This is not the opinion of security professionals.

Wayne Gilbert, Assistant Director in charge of the FBI Intelligence Division stated in May 1992 that "We're seeing very, very little reduction in total activities, counting both the SVR (Russian Foreign Intelligence Agency) and the GRU (the military counterpart). They're running at almost the same amount of activity as they have in the last two years. That's pre-coup and post-coup" ("Russian Spy Services," 1992).

In addition, espionage of the 1990s is increasingly turning to the economic side. Spy services of both our traditional adversaries and friends alike are heightening their efforts to achieve economic advantage. The days when nations focused entirely on military conquest are being replaced by efforts at economic domination.

As a counterpoint to the espionage database, PERSEREC is building a database of illegal technology transfer cases and other instances of economic espionage.

References

Russian Spy Services are Continuing Aggressive Activities, FBI Says. (May, 1992). Advisory. Washington, DC: National Security Institute.

DOD Security Review Commission. (1985). Keeping the nation's secrets: A report to the Secretary of Defense by the Commission to Review DoD Security Policies and Practices. Washington, DC: Office of the Secretary of Defense.

ESPIONAGE AND BETRAYAL OF TRUST

Joseph P. Parker
Martin F. Wiskoff
BDM International, Inc.

Problem and Background

Despite the dismantling of communist governments in the former Soviet Union and Eastern Europe, the threat to our national security remains, and efforts to obtain valuable U.S. secrets by many nations, even those thought to be friendly, will continue.

One of the most important elements in the national security process is the screening of personnel entering sensitive or high-security occupations. The existing personnel security system relies heavily on background information, such as law violations, financial problems, and drug or alcohol abuse, in making decisions about the suitability of individuals for such positions. This system has evolved, however, without any clear understanding of those individual and situational factors that may predispose certain cleared individuals to commit espionage. In particular, the theoretical framework for approaching this issue is lacking.

In response to this lack of research, the present study attempts to: (a) define the temperament/personality constructs that could be related to the act of trust betrayal; and (b) identify a potential set of existing instruments for measuring these constructs. These objectives were accomplished by conducting a literature review and consulting with experts from various academic areas.

Definition of Terms

The subject of trust, its conceptualization, and definition have been dealt with extensively in the literature of twentieth century social sciences. The term *trust betrayal*, however, does not represent a recognized area of research. Nonetheless, the behaviors of which trust betrayal forms an integral part (embezzlement, fraud, infidelity, etc.) are numerous and many have been studied in psychological, sociological, and criminological terms.

The expression *trust betrayal* conveys a number of important ideas in relating the term to espionage. The term implies the existence of a bond of trust between an individual and a larger entity (workplace organization, nation) prior to the act of betrayal. In addition, a demonstration of trustworthiness on the part of the individual is generally required before such a trust can be conferred. In this sense, then, betrayal of trust between individuals with no mutual bonds is conceptually distinct from our definition.

Literatures Reviewed

The trust literature is dominated by sociological and organizational approaches which emphasize the importance of environmental factors and regard explanations of trust betrayal at a purely individual level as incomplete. Studies using trait measurement instruments have detailed some constructs that may be related to trust betrayal at an interpersonal level, but this model was found inappropriate for explanations of trust violations such as espionage which occur in an occupational setting. The gaming approaches favored by behaviorists for exploring trust and betrayal were also deemed inadequate for the same reason.

The white-collar crime literature provided a wealth of research and theory that was found particularly suitable for describing offenses involving major violations of trust. Studies in white-collar crime using psychometric instruments provided some empirical evidence

for the use of such tests in identifying potential trust betrayers within organizations. The many empirical studies and reviews of integrity testing provided detailed accounts of the constructs related to employee delinquency. The success of biographical data in tapping some of these constructs has also been noted in recent studies.

Most of the literature directly related to espionage is either historical, fictional, biographical, or political in nature. Nonetheless, a small body of academic literature characterized by two distinct approaches was discovered: (a) attempts to explain espionage from a sociological perspective that alluded to the importance of the offender's rationalization process and described situational influences on espionage that may be more important than factors of personality; and (b) psychometric approaches that focused on the profiling of personalities that may be at risk for betrayal, using motivation to differentiate among major types of spies.

Findings

Trait and Temperament Constructs Related to Espionage and Other Offenses

Table 1 presents the individual personality traits that have been tentatively identified by researchers as predisposing to espionage, white-collar crimes, and occupational deviance and delinquency. Most, but not all of the character dimensions listed in Table 1 represent measurable trait constructs.

Regarding espionage, a wide range of motivations have been proposed, but money, ideology, and resentment/revenge have figured most prominently. There has been very limited theoretical and empirical research concerning the last two motivations, so only the money motivation will be considered in the following discussion. Pecuniary gain is assumed to be the primary motivator for offenses other than espionage, since most involve property theft.

Espionage. Authors such as Sarbin (1991), Scheibe (in press), and others have remarked on the futility of searching for a personality configuration predictive of spying, but have nevertheless offered some personality characterizations of spies. Sarbin's work is a criminological theory of espionage that divides offenders into egoistic and ideological subtypes, while Scheibe describes how self-control and social control determine an individual's susceptibility to the lures of espionage. Sarbin sees the experience of alienation or low esteem, as well as a tendency to risk-taking, as contributing conditions. Scheibe is in general agreement, but he emphasizes the lack of individual self-control.

Psychometric approaches to espionage have focused on developing personality profiles to help in identifying people who might be at risk for betrayal. Gough (1991) provides some empirical data to support use of the CPI (California Psychological Inventory) in identifying individuals with personality syndromes that might be predisposing to trust betrayal. He presents one theoretical model named Self-Centered and Self-Indulgent (John Walker Syndrome) and provides an experimental design for validation of the syndrome. Preliminary results based on this design showed that the CPI scales Self-Control, Responsibility, and Achievement Via Conformance discriminated between interpersonal trust violators at statistically significant levels.

Hogan and Jones (in press) identify a personality type called the mercenary trust violator, which they feel can be identified using psychometric instruments such as the Hogan Personality Inventory (HPI) and the Inventory of Personality Disorders (IPD). Mercenary trust violators are likened to blue-collar workers in high-risk occupations, who take large risks for relatively small reward and must demonstrate competence in the performance of their jobs over a period of

Table 1

Personality Traits Identified with Espionage and Other Offenses

Espionage

Authors(s)	Instrument(s)	Specific Trait(s)
Sarbin (1991)	No instrument	sensation-seeking; self-reliant and autonomous; alienation or low self-esteem
Scheibe (in press)	Reference made to CPI	boredom, freedom from conscience, alienation, disaffected, thrill-seeking, (-) Self-Control
Gough (1991)	California Psychological Inventory (CPI)	(-)Self-Control[1], Responsibility, Achievement Via Conformance
Hogan & Jones (in press)	Hogan Personality Inventory (HPI), Inventory of Personality Disorders	(+)Intellectance[2], Sociability, Antisocial, Paranoia, Adjustment, (-)Prudence
Community Research Center (1992)	16PF, MMPI, MCMI, IPI, SILS, NEO	(+) Psychopathic Deviate, Narcissism, Hysteria

White-Collar Offenses

Author(s)	Offense(s)	Instrument(s)	Specific Trait(s)
Gottfredson & Hirschi (1990)	All offenses, including white-collar crimes	No instrument specified	low self-control
Welch (1990)	Embezzlement	Rokeach Values Survey	(+)Self-Centered
Collins (1991)	White-collar crimes	CPI, PDI Inc. employment inventory, Biographical Ques.	(-)Socialization, Responsibility, Tolerance, Self-Control, Performance (+) Extracurricular Activity

Occupational Deviance and Delinquency

Frost & Rafilson (1989)	Employee theft, counter-productivity, drug abuse	Personnel Reaction Blank, Personnel Selection Inventory	(-)Dependability, Conscientiousness, Impulse control
Kochkin (1987)	Employee theft	16PF, Reid Report	(+)Apprehensive, Tense, Anxiety (-)Ego Strength, Conscientiousness, Controlled
Moore & Stewart (1989)	Employee theft	16PF, PSI, Reid Report, PIIP	(+)Assertive, Tense (-)Conscientiousness, Controlled, Tough-Minded
Gough & Bradley (1992)	Delinquency	CPI	(-)Socialization, Responsibility, Tolerance, Ego Integration
Hogan & Hogan (1989)	On-the-job employee delinquency	HPI Employee Reliability Index	(+)Experience seeking, Enjoys crowds, Exhibitionistic, Depressed, Guilt (-)Easy to live with, School success, Avoids trouble, Sense of attachment

1 Dash indicates inverse relationship between following constructs and criterion.

2 Plus indicates positive relationship between following constructs and criterion.

years. They point to their research with these occupational groups (law enforcement, bomb disposal, etc.) in which these workers have been found to be bright, self-confident, tough, exhibitionistic, and non-conforming.

Individuals actually convicted of espionage were given a battery of psychological tests. The most striking result of the research was that nearly 80% of the spies showed elevated scores on the Psychopathic Deviate scale of the Minnesota Multiphasic Personality Inventory (MMPI). High scorers on this measure have been described as self-centered, impulsive, likeable, exhibitionistic, alienated, and antisocial. Several convicted spies also showed raised scores on measures of Narcissism and Hysteria.

White-Collar Crimes. The white-collar crime literature shows few attempts, with one or two recent exceptions (see Collins 1991; Welch, 1990), to relate empirically personality traits with the commission of crime. Qualitative approaches to describing the personalities of white-collar offenders have been more common. Most of these have been small case studies that have produced interesting and enlightening personality profiles of white-collar criminals and have focused on the offenders' a priori justifications for committing the crimes.

Gottfredson and Hirschi (1990), in their general theory of crime, argue that low self-control, or a tendency to seek self-gratification without concern for others, is the only trait that distinguishes between white-collar offenders and the law-abiding populace.

In a study by Welch (1990), the Rokeach Value Survey (RVS) was used to assess the value systems of incarcerated embezzlers and other criminals at a minimum security correctional facility. One significant finding was that, when compared to the general population, those incarcerated were more self-centered, and placed "a lower importance on those values which do not have immediate or personal relevance" (p. 159).

A recent study (Collins, 1991) using temperament, biodata, and integrity instruments found these measures successful in discriminating between trust violators and a control group. The instruments were administered to 365 incarcerated white-collar criminals and 344 white-collar employees holding positions of authority. Analyses resulted in the creation of a six-factor discriminant model. The CPI scales Socialization, Responsibility, Tolerance, and Self-Control entered into the model, along with an integrity/honesty scale and a biodata measure of extraversion.

Occupational Deviance and Delinquency. Empirical studies of occupational deviance and delinquency can be divided into two general areas: studies of honesty/integrity evaluation used to screen employees and applied theories of delinquency that rely on trait/temperament instruments. The rationale for grouping these areas together is that many integrity tests currently purport to predict global performance and forms of counterproductivity other than theft and the term *delinquency* would include employee theft.

Several recent studies have focused on the predictive validity of popular honesty/integrity tests. While there is some disagreement, the evidence for their validity appears convincing. However, the content of these tests is not well understood. Table 1 presents three studies (Kochkin, 1987; Frost & Rafilson, 1989; Moore & Stewart, 1989) that explored the content of some popular integrity tests. Overall, agreement exists on their measurement of the constructs conscientiousness, internal control, and tension.

Psychological and sociological studies in delinquency and crime have provided a great deal of evidence concerning the personality traits that contribute to deviance.

Gough and Bradley (1992) reviewed much of the delinquent and criminal behavior

studies that have used the CPI and provide empirical support for the revised CPI's validity in predicting delinquency. They found that for the revised CPI the scales that differentiate best between delinquents and nondelinquents were Socialization, Responsibility, Tolerance, and Ego Integration. The Socialization scale has been recognized by many researchers as one of the best-validated and most powerful personality scales available for differentiating delinquent from nondelinquent groups.

Hogan and Hogan (1989) have built on research concerning the CPI Socialization scale and fashioned an employee reliability measure using the HPI. Its four main themes are: (a) hostility to rules, (b) thrill-seeking impulsiveness, (c) social insensitivity, and (d) alienation. They have found that scores on this measure relate to a wide array of positive and negative work performance behaviors.

In comparing the constructs posited for measuring tendencies to trust betrayal (espionage, white-collar crime) and delinquency a few observations can be made. In all, we find three traits have been mentioned in more than one study as predisposing to espionage: (a) poor self-control (b) thrill or sensation-seeking (risk-taking, a component of the HPI Prudence scale is included), and (c) a sense of alienation.

In the white-collar crime studies presented, we find that only self-control is consistently mentioned as a contributing factor. When studies in employee theft and delinquency are considered together, we observe that once again internal control is consistently noted. In addition, irresponsibility and lack of conscientiousness, along with elevated tension and anxiety, contribute to the commission of a wide variety of offenses.

It has been noted that most trust violators would have to establish themselves as dependable, conscientious workers before being promoted to a position of trust, so these individuals could not be grossly irresponsible or unreliable in their behavior. Therefore, we would not expect to see depressed scores on measures of conscientiousness or responsibility, as was seen with the occupational deviance and delinquency criteria. By the same token, exceedingly anxious or tense individuals would not be seen as emotionally fit for occupations requiring high-level clearances or access to large amounts of money. The inclusion of these constructs seems to indicate that they do not identify the trust betrayal criterion.

A total lack of self-control might prevent a person from obtaining a position of trust, but an overwhelming concern with self-gratification, accompanied by a tendency to follow momentary impulses, would not necessarily be disqualifying. Similarly, a tendency to take risks could not be used as a disqualifying factor and may actually be favored in certain high security occupations. Alienation from coworkers and from patriotic sentiments are of course undesirable attributes. The measurement of self-control and thrill-seeking traits is supported by a considerable body of research, but identifying alienation with paper-and-pencil tests presents several problems.

Conclusion

The constructs identified in this review provide a foundation for confirmatory analyses with other groups of trust violators. An experimental battery of tests such as that used by Collins (1991) including the CPI, an honesty/integrity test, and a biographical inventory could be administered to applicants to high-security occupations in the government, for example. However, cautionary advice to such a narrow individual approach has been given by several sociologists and organizational theorists.

Sarbin and others have shown that most spies and white-collar criminals did not obtain their positions with the intent to betray the trust bestowed on them. He reasons, therefore, that the causes of espionage will probably not be found by examining an

individual's pre-employment background. Sarbin concludes that the causes of espionage will more likely be found in the social contexts and conditions that prevail at the time of the incident. In fact, recent research (Wood & Wiskoff, 1992) on Americans committing espionage on the U.S. since World War II has documented that a contributing factor in some spying is associated with job and situational variables. Therefore, any psychometric approach to trust betrayal should also attempt to account for situational/organizational factors in empirical designs or risk being incomplete.

References

Collins, J. M. (1991). *White collar criminality: A prediction model.* Unpublished doctoral dissertation. Iowa State University, Ames.

Community Research Center (1992). *Personality characteristics of convicted espionage offenders: A slammer psychology team technical report.* Newington, VA. Author.

Frost, A. G., & Rafilson, F. M. (1989). Overt integrity tests versus personality-based measures of delinquency: An empirical comparison. *Journal of Business and Psychology*, *3*, 269-277.

Gottfredson, M. R., & Hirschi, T. (1990). *A general theory of crime*. Stanford, CA: Stanford University Press.

Gough, H. G. (1991). Personal Communication. April 16, 1991.

Gough, H. G., & Bradley, P. (1992). Delinquent and criminal behavior as assessed by the revised California Psychological Inventory. *Journal of Clinical Psychology*.

Hogan, R., & Hogan, J. (1989). How to measure employee reliability. *Journal of Applied Psychology*, *74*, 273-279.

Hogan, R., & Jones, W. H. (in press). The psychology of trust violation. In T. R. Sarbin (Ed.), *Trust and betrayal: Studies in citizen espionage.* New York: Springer Verlag.

Kochkin, S. (1987). Personality correlates of a measure of honesty. *Journal of Business and Psychology*, *1*, 236-247.

Moore, R. W., & Stewart, R. M. (1989). Evaluating employee integrity: Moral and methodological problems. *Employee Responsibilities and Rights Journal*, *2*, 203-215.

Sarbin, T. R. (1991, July). *Domestic espionage: A criminological theory.* Paper presented at Temperament Constructs Related to Betrayal of Trust Seminar, Defense Personnel Security Research and Education Center, Monterey, CA.

Scheibe, K. E. (in press). The temptations of espionage: Self control and social control. In T. R. Sarbin (Ed.), *Trust and betrayal: Studies in citizen espionage.* New York: Springer Verlag.

Welch, M. (1990). The value systems of incarcerated embezzlers compared to other inmates and the general population. *Journal of Offender Counseling, Services and Rehabilitation*, *15*, 155-164.

Wood, S., & Wiskoff, M. F. (1992). *Americans who spied against their country since World War II.* (Technical Report PERS-TR-92-005). Monterey, CA: Defense Personnel Security Research and Education Center (For Official Use Only).

ASSESSING THE ESPIONAGE VULNERABILITY OF POSITIONS

Kent S. Crawford
Defense Personnel Security
Research Center (PERSEREC)

Michael J. Bosshardt
Personnel Decisions Research
Institutes, Inc.

BACKGROUND

The continuing evaluation of cleared personnel is a critical component of the Department of Defense (DoD) personnel security program. In a recent study, PERSEREC examined the effectiveness of continuing evaluation programs in the military services (Bosshardt, DuBois & Crawford, 1991; Bosshardt, DuBois, Crawford & McGuire, 1991). This study included a survey of 60 Army, Air Force, Navy and Marine Corps installations around the world and resulted in over 50 recommendations for improving continuing evaluation.

One key recommendation focused on targeting more resources toward those at greatest personnel security risk. Specifically, security managers indicated that more continuing evaluation emphasis should be devoted to individuals who are in certain positions and geographical areas. However, only 45 percent of the sites surveyed targeted more continuing evaluation resources toward certain positions. Furthermore, nearly all those targeting efforts were based entirely on level of access (e.g., Top Secret vs. Secret) rather than on specific position factors. In other words, persons within a given level of access are treated similarly from a continuing evaluation standpoint, despite obvious differences in expected personnel security risk for different types of positions.

Other factors also point to the need for targeting some continuing evaluation resources toward high risk personnel. The large number of cleared personnel and the amount of classified information make it increasingly difficult to meet continuing evaluation objectives during the significant downsizing and resource reductions occurring during the 1990s. For example, in the 48 collateral sites surveyed (i.e., sites with personnel who have Top Secret or Secret access), there was an average of only one security manager at the installation level who was devoted to continuing evaluation for every 10,000 cleared personnel. Overall, these findings suggest that allocating resources based on personnel security risk may be an excellent approach for improving or maintaining personnel security during a decade of shrinking resources.

Few attempts have been made to conduct studies aimed at developing procedures to target continuing evaluation resources based on risk assessments. One exception was a pilot study by Abbott and Rosenthal (1990) which resulted in a preliminary instrument for assessing position vulnerability. Using a panel of subject matter experts, these authors identified nine position characteristics (e.g., sustained accessibility of foreign agents to position incumbent, routine access to sensitive information, sensitivity of information) that were hypothesized to relate to espionage susceptibility. Experts then rated a sample of positions on each characteristic. For each position, the position characteristics

were weighted and then summed to create an overall position vulnerability (PV) score. One limitation of the study was that the experts lacked extensive knowledge of counterintelligence and recruitment of spies.

The purpose of the present study was to systematically develop a more refined and comprehensive PV form using counterintelligence personnel as subject matter experts. These personnel would have extensive knowledge of how foreign intelligence officers target our cleared personnel. Abbott and Rosenthal's findings were used as the starting point for the study.

DEVELOPMENT OF A POSITION VULNERABILITY FORM

Development of the position vulnerability form involved three steps: (1) identifying a set of position vulnerability factors, (2) developing rating scales to measure these factors, (3) developing scoring procedures for the position vulnerability form. Each of these steps will be briefly described.

Identification of Factors

The initial step in the development of a position vulnerability form was to identify position characteristics or factors that are related to susceptibility to espionage. We began this process by reviewing available information and materials on espionage cases and espionage prevention. This review process resulted in the examination of over 100 reports, security manuals, and newspaper articles on espionage-related cases, as well as variables in the PERSEREC espionage database. More than one hundred possible position vulnerability factors were identified.

We reduced this large number of possible position vulnerability factors using a two-step procedure. First, we eliminated factors that emphasized personal characteristics (e.g., frequency of substance abuse) rather than position characteristics (e.g., availability of employee assistance programs). The remaining factors were then sorted into more general categories on the basis of similar content. This process resulted in 17 preliminary position vulnerability factors.

After identifying a preliminary set of position vulnerability factors, we prepared a questionnaire to obtain information about these and other possible factors. More specifically, this questionnaire asked respondents to revise the wording of these preliminary factors, identify additional factors, and combine related factors. Forty-six counter-intelligence personnel from nine government agencies with intelligence missions returned completed questionnaires. On the basis of their comments, we revised several factors, combined two factors into a single factor, and added three new factors. The resulting list contained 19 factors.

We then met with 11 counter-intelligence experts from the same nine government agencies to further refine these factors. During this workshop, participants revised the existing position vulnerability factors, eliminated two factors, and added one new factor. Following this, participants organized the final 18-position vulnerability factors into four general categories: (1) access and exposure to classified/sensitive information, (2) job factors, (3) threats from potential

contacts, (4) security countermeasures. Table 1 shows the factors included in each category.

Development of Rating Scales

After identifying and defining a set of position vulnerability factors, we developed rating scales to measure these factors. The first step in the rating scale development process was to obtain examples that describe very high, moderate, and low security risk levels for each position factor. This was accomplished using a questionnaire survey. Specifically, for each position factor, questionnaire respondents wrote one or two short examples that described a position at very high risk from a personnel security standpoint, one or two examples that described a position at moderate risk, and one or two examples that described a position at very low risk.

Fifteen counterintelligence personnel from seven government agencies with intelligence missions returned completed questionnaires. These questionnaires contained nearly two thousand possible rating scale anchors.

Because of the large number of examples, it was necessary to develop rating scale anchors that summarized the content of several examples. To accomplish this, we sorted these anchors according to factor, scale level (very high, moderate, or very low security risk), and similarity in content. Five anchors (very high, high, moderate, low, or very low security risk) were then written for each position vulnerability factor. This process resulted in a preliminary set of 18 rating scales, one scale per position vulnerability factor, each scale defined at five levels with specific examples.

After developing a preliminary set of position vulnerability rating scales, we met with seven counterintelligence experts from seven government agencies to revise the position vulnerability rating scales and evaluate the importance of each position vulnerability factor. In addition, the experts rated the absolute importance of each position vulnerability factor in terms of its importance in determining the vulnerability of a position to espionage. They did this by allocating 100 points across the 18 factors. Table 2 presents the results. Factors receiving the highest mean importance ratings are sensitivity of classified/sensitive information, contact with foreign nationals, job location threat, and frequency of access to classified/sensitive information. Factors with the lowest mean ratings are position stress and employee assistance programs.

Development of Scoring Procedures

We used the mean factor importance weights as a basis for weighting the importance of each position vulnerability factor. The actual rating for each factor (which is made on a 1 to 5 scale) is multiplied by the factor weight. A score for a given position is computed by summing the weighted factor scores for each of the 18 factors. The mean factor importance weights have been transformed so that the total score on the position vulnerability can range from 0 to 100.

APPLICATIONS

The position vulnerability form has several possible applications. These include:

1. Determining the extent of initial background screening. Information from the position vulnerability form could be used to identify which positions are most

Table 1
Position Vulnerability Categories and Factors

A. Access and Exposure to Classified/Sensitive Information
   1. Frequency of Access to Classified/Sensitive Information
   2. Range/Amount of Access to Classified/Sensitive Information
   3. Sensitivity of Classified/Sensitive Information
   4. Potential for Unauthorized Exposure to Classified/Sensitive Information
B. Job Factors
   5. Career Field
   6. Position Oversight
   7. Position Status
   8. Position Stress
   9. Cost of Living/Compensation Pressures
C. Threats from Potential Contacts
   10. Contact With Foreign Nationals
   11. Job Location Threat
   12. TDY Travel
D. Security Countermeasures
   13. Information Security Safeguards and Procedures
   14. Physical Security Safeguards and Procedures
   15. Personnel Security Procedures
   16. Security Education Procedures
   17. Employee Assistance Programs
   18. Availability of Support Systems

Table 2
Ranking of Position Vulnerability Factors
Based on Mean Importance Ratings

Mean Importance Rating	Position Vulnerability Factor
9.59	Sensitivity of Classified/Sensitive Information
8.54	Contact With Foreign Nationals
7.73	Job Location Threat
7.73	Frequency of Access to Classified/Sensitive
7.38	Range/Amount of Classified/Sensitive
7.38	Career Field
6.79	Personnel Security Procedures
5.50	Information Security Safeguards and Procedures
5.04	Cost of Living/Compensation Pressures
4.45	Physical Security Safeguards and Procedures
4.33	TDY Travel
4.22	Position Oversight
4.22	Security Education Procedures
4.10	Position Status
3.62	Potential for Unauthorized Exposure to Classified/Sensitive Information
3.28	Availability of Support Systems
3.05	Position Stress
3.05	Employee Assistance Programs

vulnerable from a security risk stand-point. This information could in turn be used to determine the extent of an initial background screening. Persons who are being considered for positions that are most vulnerable would receive more extensive screening. In contrast, persons under consideration for positions that have low position risk would receive less extensive screening.

2. Prioritizing positions for follow-up reinvestigations. As noted in the introduction, a continuing evaluation program must operate in an environment of limited resources and continual pressure to reduce expenditures. Information from the position vulnerability form could be used to identify which positions are most vulnerability from a security risk stand-point. This information could in turn be used to prioritize positions for follow-up reinvestigations. Position holders in positions that are most vulnerable would receive more frequent reinvestigations. Position holders in positions that have low position risk might be reinvestigated less often.

3. Providing input for security education and security awareness. Information about the security risk associated with various cleared positions could be provided to persons responsible for security education and awareness programs. Such information would help these individuals tailor security programs to reduce risks associated with position vulnerabilities. For example, persons occupying positions involving frequent contact with foreign nationals would receive additional security emphasis on this vulnerability area. The position vulnerability form information would provide a means of identifying which position vulnerabilities are most salient for a job, a unit, or the installation as a whole.

4. Use of security countermeasures. The position vulnerability form could be used to identify groups of individuals or organizational units that have a higher than average vulnerability. More strict security countermeasures (e.g., information and physical security) and other approaches (e.g., increased oversight, better employee assistance programs) could be implemented to reduce the vulnerability of the group or organizational unit.

5. Assignment of cleared personnel to positions. Position vulnerability information could be combined with personal vulnerability variables to minimize security risk during the assignment of cleared individuals to minimize risk. Within a given agency, some positions will probably be more vulnerable from a security risk standpoint than other positions. Job incumbents with higher personal vulnerability risk profiles would be placed into positions with less security risk. Conversely, incumbents at less risk according to personal vulnerability standpoint might be placed into positions at greater risk from a position vulnerability standpoint.

6. Use in analyzing espionage cases. The position vulnerability form could provide valuable insights into the determinants of espionage. Position vulnerability information for known spies could be compared with position vulnerability information for non-spy populations. The results of the study would provide information about the factors associated with espionage. This information, in turn, would be used to modify all personnel security procedures

(e.g., initial investigations, reinvestigations, security education) to reduce the threat of espionage.

FUTURE RESEARCH

This study was the first phase of a longer term initiative to develop and implement improved personnel security procedures based on application of position vulnerability assessment. PERSEREC is currently developing a computerized version of the position vulnerability form. This adaption will allow greater flexibility and ease in making and recording assessments in the field.

Once the PC-based form is developed, the next step is to pilot test the automated instrument within a single organization. This information will be used to evaluate the adequacy of the rating form instructions and scales as well as the feasibility of the rating process and scoring system. Results of the pilot test will provide information for refining the computerized position vulnerability form prior to actual implementation.

REFERENCES

Abbott, P. S. and Rosenthal, D. B. (1990). *Development of the Position Vulnerability Assessment (PVA) instrument* (Draft Final Report-90-02). Alexandria, VA: HumRRO International, Inc.

Bosshardt, M.J., DuBois, D.A. and Crawford, K.S. (1991). *Continuing assessment of cleared personnel in the military services: Report 4 - System issues and program effectiveness* (PERS-TR-91-004). Monterey, CA: Defense Personnel Security Research and Education Center.

Bosshardt, M.J., DuBois, D.A., Crawford, K.S. and McGuire, D. (1991). *Continuing assessment of cleared personnel in the military services: Report 2 -Methodology, analysis, and results* (PERS-TR-91-002). Monterey, CA: Defense Personnel Security Research and Education Center.

Getting Serious About Security Awareness: Better Briefings, Videos, and Posters Alone Won't Do It

James A. Riedel
Defense Personnel Security Research Center
Monterey, CA 93940-2481

Joseph P. Parker
BDM International, Inc.
Monterey, CA 93940

Preventing the betrayal of trust, more specifically the disclosure of secrets by insiders, is crucial to the success of organizations that must protect secrets. There are three approaches to control the disclosure of secrets: develop information and physical security safeguards, select personnel for positions of trust, and develop a program of deterrence (Sarbin, 1992).

This research focused on the Defense Department's program of deterrence that attempts to prevent espionage through security awareness education. These educational activities are intended to teach personnel how to protect secrets and to motivate their compliance with security procedures.

Security awareness education in the Defense department is driven by requirements that flow from Presidential Executive Orders to agency or departmental regulations. Overall, the Service and agency requirements for security education consist of five different types of briefings: initial, refresher, foreign travel, counterintelligence, and termination. In addition, local regulations concerning security education in the form of training manuals, pamphlets, supplements, handbooks, etc. form an important part of security education at many installations.

Despite the recent and dramatic reductions in tensions between East and West and the decreasing military threats to our national security, security awareness is as important as ever. There has been little reduction in intelligence gathering against the US by adversaries ("Russian Spy Services", 1992). The dire economic conditions in Eastern Europe and the former Soviet Union have increased pressure on these countries to gather crucial science and technology information. Additionally, the need for economic information has led to an increase in economic espionage against the US by many non-traditional spies, including the intelligence services of our allies.

At the same time circumstances at home are creating a fertile ground for foreign intelligence operatives. Our own recession and military downsizing have resulted in unemployed defense workers who might be disgruntled or in desperate financial straits. Workers predisposed to betray their country might be tempted to trade critical information for revenge or money. Previous research has found financial problems and revenge to be related to espionage (Wood & Wiskoff, 1992).

The need to improve security awareness education surfaced in a 1985 report from the Stilwell Commission (Department of Defense (DoD) Security Review Commission). The Commission suggested that DoD could avoid some duplication of effort and improve the quality of briefings and training aids by better coordinating and facilitating its programs. In October 1988 the House of

Representatives published *U.S. Counterintelligence and Security Concerns: A Status Report. Personnel and Information Security*. This report, produced by the Subcommittee on Oversight and Evaluation of the Permanent Select Committee on Intelligence, concluded that "Not enough is done to promote security awareness." Likewise, more effective security awareness education was recommended as a means of improving security continuing evaluation programs (Bosshardt, DuBois, and Crawford, 1991a).

While strong agreement appears to exist on the essential contribution of security awareness programs to national security, no systematic effort has been made to determine the major needs and problems of these programs. The present study addresses this deficiency by assessing the current state of security awareness education at military installations and offering recommendations for their improvement.

Method

Planning for this survey project was initiated by meetings with Service headquarters representatives and installation security staff members involved in information and personnel security. Information concerning security awareness programs and recommendations for their improvement was obtained in these meetings.

On the basis of these discussions, two survey forms were constructed. The first was a detailed interview protocol which contained a mix of closed questions (*e.g.*, yes/no, multiple choice, and rating items) and open-ended questions. The second form was a 100-item questionnaire comprised almost entirely of rating items. The content areas covered in these forms are presented in Table 1.

Table 1

Content Areas Covered in Survey Forms

<u>Respondent information</u> - pay grade, years of experience in security, and position tenure.

<u>Organizational and policy information</u> - the size and primary mission of installation; size of security office; amount of time spent by security staff on security education activities; Department of Defense (DoD), Service, and local policies governing security education.

<u>Security education practices</u> - the time spent covering various security topics and disciplines, staff expertise in specific subject matter and disciplines, training and education objectives, use of media and training methods, and sources for education media and products.

<u>Program management</u> - internal and external support for security education implementation, accountability, security awareness in performance appraisals, security inspections, security office controls, program emphasis at different organizational levels, security staff training and development, and effectiveness of security education programs.

Data were collected between July and October 1990 at a total of 58 sites (18 Army, 12 Navy, 23 Air Force, 4 Marine Corps and one DoD). At each site a researcher interviewed the installation security office representative for approximately 2 1/2 hours using the structured survey form. The 100-item questionnaire was also completed by the interviewee. Meetings were also held with small numbers of unit security representatives and/or security staff at which time only the second survey form was completed.

Survey data were received from a total of 111 individuals. Forty-seven security office representatives completed the interview form and all but seven of these also completed the questionnaire. Sixty-four unit security representatives--mostly unit security managers--completed only the questionnaire. A total of 104 questionnaire forms and 47 interview forms were completed.

Results

Security Awareness Program Effectiveness

Overall, security managers rated their security awareness programs as moderately successful. While several raters considered their programs very effective, relatively few classified them as ineffective. Respondents felt that their programs were especially effective in establishing close personal contact with installation personnel, in providing staff assistance visits and program reviews, and in distributing security reminders and other written materials on a continuing basis.

Respondents were asked to assess the current and potential effectiveness of 15 different components of a security awareness program using a 10-point scale with a low ranking of very ineffective and a high ranking of very effective. The difference between the mean ratings of potential and current effectiveness were calculated to determine how much interviewees felt each component could be improved. The availability of media products and services was the area in which the most room for improvement was noted; respondents' average scores indicated that effectiveness in that area could be nearly doubled. Other areas with room for improvement included security staff training in security awareness and the emphasis placed on security awareness.

Availability and Quality of Media Products

Security professionals repeatedly expressed concern with the availability and quality of media products. Many had virtually no access to information concerning what materials might be available and how to procure them. In addition, cost frequently prohibited them from obtaining some of the products. Lack of a reliable, timely, and sufficiently comprehensive distribution system also prevented them from acquiring more commonly available educational materials and publications.

Of the media products used in security education, much of the criticism was reserved for videotapes and movies. They were frequently faulted for their lack of relevance to local security conditions, for being out of date, and for being boring. Security managers expressed frustration in not having the available resources to develop media to meet their own specific needs.

Security Manager Training

Newly assigned security managers lack appropriate experience or training in their positions. Deficiencies were due, in part, to the nature of the career path for security managers. It provides limited opportunity for developing the knowledge and skills required to design and implement an effective security awareness program. Also, opportunities to attend training courses on the job are very limited (Bosshardt, DuBois & Crawford, 1991b). Training opportunities are not readily accessible due to their location, limited class sizes and time requirements. Shortcomings attributed to inadequate training included a lack of expertise in creating media products, in preparing and delivering briefings, in clearly articulating security threats, and in instructing personnel in technologically sophisticated areas such as computer and communications security.

Security Manager Support

Command support and emphasis was seen as essential by security managers but non-existent for half of those interviewed. In particular, few commanders or others in top leadership positions were visible in security awareness training activities, nor did they provide effective mechanisms for enabling the security manager to enforce compliance with security requirements. In addition, as might be expected, security offices perceived a shortage of budgetary and personnel resources provided by the command.

Discussion

This research suggests a need for enhanced instructional media to improve the quality of security education. There are two ways in which better security education materials could be provided to security managers in a timely fashion. First, a centralized distribution system for these materials could be implemented, making materials more accessible. Second, the quality of materials developed at DoD and Service headquarters levels could be improved. With either approach, however, there is a need to evaluate training to ensure it is clearly tied to important work place behavior that leads to important organizational results (Brinkerhoff, 1990).

But the findings also suggest that aspects of the organizational environment are undermining security awareness. For example, security managers lack the appropriate experience or training for their positions and there is a lack of command support and emphasis for security awareness programs. It may be fruitful to employ a number of organizational change strategies, in addition to education and training, to promote and maintain security awareness.

Training should be brought to the security manager rather than requiring the manager to attend formal courses at another location. The absence of career development, the fact that security educator is a part-time position and a lack of preparatory job training all suggest a low priority is given to security awareness. This could be partially ameliorated through correspondence courses, mobile training teams, or regional training that is accessible to security managers.

The lack of command support and emphasis for security awareness programs suggests a latent message--compliance with security procedures is not really very

important. The absence of visible top leadership involvement and the shortage of budgetary and personnel resources for security awareness confirm this message. This message contradicts the overt meaning of exhortations repeatedly delivered in security training--compliance with security procedures is of paramount importance. The discordance in these messages undermines the potential deterrent affect of training.

If we are really serious about understanding and maintaining security awareness we must begin to look beyond training and education alone for interventions. The exclusive use of training and education to deter the disclosure of secrets is based on conceptualizations which are too individualistic or "psychological" in focus. Training interventions ignore important aspects of organizational environment such as the social organization, individual attitudes, organizational climate, occupational culture, and rewards. These environmental influences may either sustain or erode security-relevant attitudes, motivation, and behavior.

For example, concepts of security are embedded in the organizational climate and what is called "working rules" found in organizations (Manning, 1992). Even though formal security rules and sanctions are necessary, they are only part of the picture. Employee attitudes toward security collectively reflect "working rules" which govern everyday routines used by individuals to accomplish work. These working rules may obviate, bend, circumvent or reinforce the formal written rules. In organizations with loose working rules about the use a access to documents, programs, and information, it is more likely that information will be stolen, bought or leaked than those with different working rules. Routine compromising of rules and procedures by means of working rules provides the basis for security breaches and is the context for security awareness. Future research should identify aspects of organizational environment such as organizational climate, determine their relationship to security awareness, and develop strategies to improve awareness.

Conclusion

There is a need to buttress educational efforts to promote security awareness with enhanced instructional media and materials. This could be accomplished by improving the quality of media products and implementing a more effective system for their distribution.

If we are serious about security awareness, however, improving briefing, videos, and posters alone will not be enough. Education and training interventions ignore important aspects of organizational environment. These environmental influences may either sustain or erode security-relevant attitudes, motivation, and behavior.

Security awareness education and training will work best when it is in concert and integrated with an organization's environment. We must begin to understand how aspects of an organization's environment either facilitate or reduce formal efforts to create and sustain security. Only then can we begin to develop strategies which, along with education and training efforts, encourage organizational change supportive of security goals.

References

Bosshardt, M.J., DuBois, D.A., & Crawford, K.S. (1991a). *Continuing assessment of cleared personnel in the military services: Report 2 - Methodology, analysis, and results.* (Technical Report PERS-TR-91-002). Monterey, CA: Defense Personnel Security Research and Education Center.

Bosshardt, M.J., DuBois, D.A., & Crawford, K.S. (1991b). *Continuing assessment of cleared personnel in the military services: Report 3 - Recommendations.* (Technical Report PERS-TR-91-003). Monterey, CA: Defense Personnel Security Research and Education Center.

Brinkerhoff, R.O. (December 1990). Security awareness program evaluation. *Proceedings of "Security Awareness in the 1990s,"* Monterey, CA.

DoD Security Review Commission. (1985). *Keeping the nation's secrets: A report to the Secretary of Defense by the Commission to Review DoD Security Policies and Practices.* Washington, DC: Office of the Secretary of Defense.

Manning, P.K. (September 1992). *Security awareness and organizational context.* Colloquium presented at the Defense Personnel Security Research Center, Monterey, CA.

Russian Spy Services are Continuing Aggressive Activities, FBI Says. (May, 1992). *Advisory.* Washington, DC: National Security Institute.

Sarbin, T.R. (May 1992). *The power of resentment: some observations on trust and betrayal with special reference to computer crime.* Paper presented at the Fifth Annual Conference of the Department of Defense Security Institute, Williamsburg, VA.

United States House of Representatives. (1988). *U.S. counterintelligence and security concerns: A status report. Personnel and information security.* Report submitted by the Subcommittee on Oversight and Evaluations of the Permanent Select Committee on Intelligence. Washington, DC: U.S. Government Printing Office.

Wood, S., & Wiskoff, M.F. (1992). *Americans who spied against their country since World War II.* (Technical Report PERS-TR-92-005). Monterey, CA: Defense Personnel Security Research and Education Center (For Official Use Only).

The Changing German Armed Forces in 1992

Friedrich W. Steege
German Ministry of Defense, P II 4
Bonn, Germany

Introduction

This contribution to the MTA '92 Conference deals with attitudes and behaviors towards the armed forces in Germany. There are two main problem areas. One concerns the restructuring process which the Bundeswehr now experiences. Another is related to the new developments which began with German unification. With this panel we want to provide information on the present situation as well as to report psychological data aimed at helping us to understand it and to better cope with it.

Topics of the panel are (1) a general overview of the situation in Germany based on behavioral data, with special reference to a large sample of draftees; (2) new developments in recruiting procedures; (3) data on the efforts to achieve equal opportunities for East and West German volunteers; (4) aptitude and motivation data concerning applicants for officer training; and (5) a review of the development of noncommissioned officer training.

We live in a still rapidly changing world and have to consider new facets of security psychology. In many parts of the world, security consciousness has changed dramatically. Determinants and conditions of security have been given wider attention

- the larger the number of people on earth has grown,
- the more critical and risky the supply of goods and services, particularly of health services, has become,
- the more directly the civilian population has become threatened by military conflicts, and
- the more articulately doubts have been expressed with respect to the legitimacy of conflicts.

Security has therefore become a comprehensive concern that is accompanied by deeply rooted existential uncertainties (cf. Steege & Fritscher, 1992). The most recent new quality of security is, as you all know, the concern that global survival be safeguarded. More and more, this is seen as the most important problem by many people all over the world. The scientist and political observer C.F. von Weizsäcker has coined the term "world domestic policy" ("Weltinnenpolitik") particularly in this context. The classical security (i.e. defense) policy has thus been given new dimensions. After the large-scale military confrontation has disappeared, anarchy and civil war within nations has become more frequent (former Yugoslavia, Georgia, Somalia, Afghanistan). The large ideological and totalitarian eastern power structure has been followed by a wave of nationalism which in many cases may be more properly called tribal egoism. Hence, warfare and peace-making missions of the United Nations have to be redefined in order to make international policy and law effective in national conflicts.

General overview of the situation in Germany

International meetings like that of the MTA serve not only the purpose of scientific exchange. They also have the important function to improve mutual understanding particularly in times of international upheaval.

The past centuries of world politics teach us that history never repeats itself. We are convinced that this is valid also for Germany. The torches of a long summer in our country

and the recent riots are very regrettable and shaming facts. Even though democracy has taken firm roots in Germany and the armed forces as well as public administration are aptly controlled by legislative bodies and the courts, we understand the conclusion of many foreign observers that the Germans still did not learn their lessons of history. Nevertheless, it is our firm belief that we do not have a situation similar to that in the early 1930s. There is neither systematic antisemitism nor *terrorist-type* right-wing radicalism. There is as yet no clear picture of the size and the organization of the radical right-wing. Our government and our Länder administrations have come to learn that the supporters of extreme right-wing ideas are a real problem. There are also some influences of these ideas on soldiers and applicants for our armed forces which we observe, and deal with, very carefully.

A closely connected problem area are the debates on offering and providing asylum. We can only repeat that this complex problem must be solved in the near future. Even we Germans hardly understand why consensus among our political parties on this crucial – and costly – issue has not been reached to this date.

The integration of the two Germanies will take place. But as in every societal change and development, after more than one generation of fundamentally different political systems, the process of growing together will take time. The main problems may in fact be the short-lived memory of the Germans and the impatience they are said to show regularly, and misleading political metaphors that created hopes that "blossoming landscapes" would be attained in a short time. It would be better if more people realized the cost of 50 years of dictatorship and suppressing regimes in the eastern part of our country rather than talked about the cost of unification almost daily .

Integration of the armed forces in the society

The role armed forces play in a society determines the self-appreciation of the members of the armed forces. The integration of the Bundeswehr in our society has been a topic since the creation of armed forces in former West Germany. The success of restructuring our armed forces and of integrating personnel of the East German Army in the Bundeswehr depends on the general role armed forces play in our country.

It has still to be stressed that the *relation of society and armed forces* in Germany is a peculiar one, and in many respects unique in the world (cf. Steege, 1991b). Our recruiting situation is very severely hampered by the fact that we grant the right of conscientious objection in a unique fashion. The number of conscientious objectors is still increasing, at best leveling off. In consequence, the recruiting of personnel has become very difficult for the Bundeswehr; there are hardly enough conscripts willing to serve and not enough volunteers willing to enlist. And our recruiting organization will be facing more serious difficulties in the near future (cf. Ebenrett, 1992).

Normalization of the Bundeswehr after unification has not yet been achieved. The intake of *former East German Army* officers and noncommissioned officers has been prepared in the meantime by the independent commission set up for this purpose (cf. Steege, 1991b). The first have received their contracts, but only a limited number can be hired. This concerns particularly officers. We have further to regard different roles that officers and noncommissioned officers played at various task levels in the former East German Army. Many officers and particularly noncommissioned officers of the Bundeswehr have a considerably higher responsibility than did their counterparts on the same rank level in the East German Army. This is regarded as a challenge by those who are given the higher responsibility. Another lasting problem are inequalities of payment. There is no satisfactory solution in view yet for the

consideration of the service time in the former East German Army in the pension scheme of the Bundeswehr.

With respect to *Bundeswehr applicants from our eastern states without prior military service,* we have found that levels of schooling have an influence. The officer applicants from the eastern states make up about 20% of the total group, matching almost exactly the distribution of our population (17 million East Germans, 63 million West Germans). Whereas 56% of the applicants from the western states are selected, only 47% of the East German applicants are successful. The rejection rate thus is about 9% higher for applicants from the eastern states. For the level of enlisted personnel I refer to the data of Rodel (1992) who found no significant differences between East and West German applicants concerning military qualification. We may cautiously assume that high school graduates from the Eastern states in 1992 have a slightly lower level of capacity. Applicants for enlisted service with intermediate-level schooling, however, match the mental requirements and the performance of their western peers.

Psychological findings concerning East and West German personnel

The unification of Germany has given rise to a number of a macro-level field studies. One area of analysis is the process of restructuring of large-scale organizations like the armed forces. As Sinaiko (1992) pointed out, psychological measures include to a considerable extent contributions to organizational change and development. The Handbook of Military Psychology covers a broad range of activities of applied psychology in this field (Gal & Mangelsdorff, 1991; cf. also Steege, 1992).

In this context, I will first provide some selected data of a poll in the German military recruiting organization conducted in March 1992 and reported to the Federal Minister of Defense in June 1992. Further, I will briefly summarize findings of Wottawa (1992) presented at the 28th Symposium of Applied Military Psychology (IAMPS) in Berlin.

Results of a poll of German conscripts in March 1992

The difficulties of recruiting personnel were the main reason to design an opinion poll covering about 6,000 draftees of a representative sample spread over Germany. The main aim was to find out how many potential volunteers would be contained in the age-group surveyed. In more detail, the social-psychological research was designed to gather data concerning the following questions: (1) Attitudes towards the German armed forces and its present and future tasks, (2) repercussions of being drafted on the social environment of the draftees, (3) evaluation of the sources of information about the Bundeswehr, (4) evaluation of the information and counselling activities during medical examination and psychological aptitude and placement testing, (5) opinion on possible reasons why young men do not apply as volunteers, (6) evaluation of measures considered promising to win volunteers, and (7) rating of the readiness of conscripts to enlist.

The young men were queried following their FAF aptitude and placement test battery in regional recruiting offices or in the officer and volunteer selection centers. The sample consisted of 4,900 draftees (DR), 570 applicants for enlisted service (ES), and 400 officer applicants (OA). Some interesting results of this research are: 96% of the OA and 95% of the ES, but only 42% of the DR regard the Bundeswehr today as "urgently necessary" or "necessary". The East and West German draftees show no considerable differences. Twenty percent of the sub-sample of OA is formed by applicants from the new states, they also do not differ from the evaluation of their western peers. With respect to the present and future tasks of the Bundeswehr we received answers as follows: The most similar answers were given concerning the external security which armed forces are expected to guarantee. More differentiated

– as expected – are the reactions when UN peace-keeping activities and especially when UN peace-making activities are in question. UN peace-keeping missions are agreed to by 43% of our DR, whereas 35% of them are not yet decided on this issue. By comparison, 66% of the ES and 71% of the OA agree. Forty-seven percent of the DR, however, reject the idea of taking part in UN combat missions, whereas the majority of the ES (55%) and the OA (57%) regard them as worth discussing. Interesting results are received with respect to possible environmental protection tasks of the Bundeswehr. Sixty-seven percent of the DR agree to this, no matter what it may mean.

Support for the Bundeswehr by acquaintances is a problem and is a likely explanation for many of the recruiting difficulties. The Bundeswehr is being rejected by the majority of the acquaintances of DR (67%), whereas ES (40%) and OA (46%) have less difficulty in this respect. The repercussions of possible enlisted service on the relationship with wife or girlfriend are particularly important. "Critical" and "negative" ratings are given here by 76% of the DR, 36% of the ES and 54% of the OA.

Comparison of East and West German private industry managers

The research of Wottawa started shortly after unification. It revealed that, psychologically, people in the two German states had developed much further apart than had originally been assumed. Marked differences evident in everyday life concern leadership qualities, initiative and assertiveness, qualities which, certainly not without exception, are found developed to a lesser degree in the eastern states than in inhabitants of the old Federal Republic. The empirical results are based on data of applicants since a random sampling was technically unfeasible. The data of more than 3,000 applicants from the eastern states were used for preselection, 400 of the applicants were tested, and 120 finally took part in an assessment center. The sample of "Westerners" was smaller. As survey instruments two personality questionnaires were used (SERVO to ascertain the personality dimensions relevant for service orientation, the CPI for the general personality structure).

Some selected psychological mechanisms which can explain the empirically proven differences between the inhabitants of the "old" and the "new" federal states are summarized in the following. *Expected differences* as consequences of the GDR system on the part of the East Germans are a greater "need for social acceptance" and "motivation to help", lower "flexibility", "sociability" and "task identification", a higher degree of "self-monitoring", less "extraversion", more "restraint", and a higher degree of "frustration tolerance" as a reaction to getting accustomed to frequent failures. *Non-expected* were the following differences between applicants from the East and the West in executive behavior: a lower degree of "permissiveness", a higher "need for achievement", and a lower degree of the dimension "performance in non-structured situations" on the part of the East Germans. The prospects for the future suggest an automatic adaptation of the respective behavior patterns over time. Wottawa (1992) maintains however that a wait-and-see approach is not acceptable, the process has to be mediated, if not accelerated.

From his findings, Wottawa (1992, section 6) derives some consequences and caveats for the armed forces. As for the "psychological" components of unification, special importance attaches to the Bundeswehr, as it is there that many young people come into intensive contact with people from former West Germany at an early time in their lives, in the role of superiors as well as peers. Wottawa concentrates on fair selection of Eastern officers and enlisted personnel, proper selection and preparation of Western officers for assignment in the new federal states, and the avoidance of mistakes in the organizational culture of the kind that can be observed in that of business enterprises.

In general with respect to fairness of selection he notes that distorting mechanisms that operate to the disadvantage of the applicants from the East should be prevented. Therefore, examination boards should pay particular attention to making sure that by preparatory schooling of the board members mistakes operating to the disadvantage of applicants from the East be minimized. He further stresses that typical mistakes in the organizational structure should be avoided. From the experience in analyzing the private sector he derives that it appears that many enterprises, from thoughtlessness or misjudgment of the psychological situation in the new federal states, exhibit modes of behavior which are damaging to long-term beneficial cooperation. Examples are frequent derogatory remarks, "jokes" about typical behavior of people of the new German states, emotional reactions to Eastern vocabulary, overemphasis on Western status symbols etc. This may lead to a vicious circle that reinforces behavior patterns of Easterners one is aiming to change.

Discussion

Measured against the economic and general situation of our society at large, one can say that the integration of members of the former East German army into the Bundeswehr has made relatively better progress than comparable processes in other sectors of our society. If one looks at the concomitant process of meeting the target of a considerable restructuring of our armed forces, the German armed forces have, in general, coped satisfactorily with what was and still is required (cf. Steege, 1991a). Generally speaking, it seems as if a hierarchically structured social system like armed forces is less susceptible to complaints against the consequences of the process of unification of our society than, for example, the private sector.

This may seem too superficial. The individual concerned or involved will have different, at least mixed, feelings. It is a fact that many of those who had an explicit involvement in the former communist system have left at the earliest possible date, in late 1989. There were a few officers and noncommissioned officers taken in who in the meantime have been discovered to have had formal relations with the GDR security service. Since they had signed a statement saying that they did not, they were fired immediately after they have been found out. Many of the former East German Army soldiers now stress the soldierly ethics and skills they had all the time, and deny significant political involvement. It seems as if we would have to live with the fact that the individual fate *as such* within a system like the former GDR cannot be made the subject of judicial appraisal. Actually, the individual is given a new chance. Whether this is really beneficial may be evaluated by those who have an insight into, and experience in, what it means to have lived "in vain" for about 40 years, without a real chance to escape the system. This is not to say that not everybody who has played an active role in that system should be put to justice as far as this is possible in a democracy governed by the rule of law. The special authority set up for this purpose (the Gauck office) has to go through this laborious task to the end.

Let me quote Wottawa (1992) who contends that "the 'psychological' reunification of Germany is likely to take many more years. However, there is some hope that the presently massive differences and the indications of alienation between the two parts of Germany can relatively soon be overcome or at least be held at a reduced level. Efforts going in this direction will be successful the earlier, the more openly the existing differences, which developed as a necessary consequence of the different political systems, are accepted. Not infrequently, though, one can observe a tendency to deny these differences (presumably with the good intention of making them disappear by denying them); yet such behavior is certainly counterproductive. The differences between two (sub-)cultures are determined not by a common or a different language but by the respective concrete conditions of life in the past – and in part still in the present. The Bundeswehr has a special key function in accelerating and

optimizing this "psychological" reunification, since here a considerable proportion of the young will have formative experiences regarding how people in East and West Germany actually treat and deal with one another." And he adds that for this reason the Bundeswehr should make intensive use of the special psychological expertise of its psychologists from both sides of the former iron curtain.

Conclusion

The actual situation in Germany is characterized by the danger of radicalism of the right and the difficult asylum problem. We are optimistic that the former two Germanies will grow together in the foreseeable future, also psychologically. The integration of the Bundeswehr will be not a problem for long. But the situation of the Bundeswehr is still contradictory to a certain extent: On the one hand, the resentment against the armed forces seems to decrease as does the opposition against international responsibilities and participation in UN peacekeeping activities. On the other hand, the number of those not ready to join the forces does not diminish. It seems necessary that the recruiting organization for draftees gets more flexible and more successful. A further means to improve the situation is higher leadership qualification (cf. Bucher, 1992). There are first signs of an increase in the number of volunteers. To meet the future force requirements we will, nevertheless, need the full commitment of everybody concerned.

References

Bucher, E.W.H. (1992). *Computer-assisted programs in the training of leadership behavior.* Paper presented at the 34th Annual Conference of the Military Testing Association. San Diego, CA.

Ebenrett, H.-J. (1992). *Meeting the German Federal Armed Forces requirements for conscripts: A current challenge.* Paper presented at the 34th Annual Conference of the Military Testing Association. San Diego, CA.

Gal, R. & Mangelsdorff, A.D. (1991). *Handbook of military psychology.* Chichester: Wiley.

Melter, A.H. (1992). *Aptitude and motivation of accepted and rejected officer applicants.* Paper presented at the 34th Annual Conference of the Military Testing Association. San Diego, CA.

Rodel, G. (1992). *Equal opportunity for East and West German volunteers: An unresolved issue.* Paper presented at the 34th Annual Conference of the Military Testing Association. San Diego, CA.

Sinaiko, H.W. (1992). Military manpower in an era of smaller forces: Some issues and opportunities for applied psychology. In M.L. Rauch (Ed.), *Proceedings of the 28th Annual Symposium of Applied Military Psychology in Berlin, June 1992.* Bonn: MOD GE.

Steege, F.W. (1991 a). German unification, the Federal Armed Forces, and the Federal Armed Forces Psychology Service. *In Proceedings of the 33rd Annual Conference of the Military Testing Association* (pp. 354-359). San Antonio: USAF Armstrong Laboratory Human Resources Directorate and USAF Occupational Measurement Squadron.

Steege, F.W. (1991 b). Armed forces and society in Germany – A complex affair. In *Proceedings of the International Seminar on the 'Military in the Service of Society and Democracy'* (pp. 226-238). Zikhron Ya'akov, Israel: The Israeli Institute for Military Studies.

Steege, F.W. (1992). IAMPS 1963 until 1992: A brief historical review. In M.L. Rauch (Ed.), *Proceedings of the 28th Annual Symposium of Applied Military Psychology in Berlin, June 1992.* Bonn: MOD GE.

Steege, F.W. & Fritscher, W. (1992). Sicherheitspsychologie (Security psychology). In A. Schorr (Ed.), *Handwörterbuch der angewandten Psychologie.* Bonn: Deutscher Psychologen Verlag.

Wottawa, H. (1992). Results of a psychological study of personality characteristics in different German settings. In M.L. Rauch (Ed.), *Proceedings of the 28th Annual Symposium of Applied Military Psychology in Berlin, June 1992.* Bonn: MOD GE.

Meeting the German Armed Forces Requirements for Conscripts: A Current Challenge

Heinz-Jürgen Ebenrett
Federal Office of Defense Administration
Bonn, Germany

The Bundeswehr has been a conscript army since its creation in 1955. Almost 50% of its personnel strength is made up of conscripts. Conscripts have been doing their duties for more than 35 years now to the satisfaction of the forces, and, within NATO, they have guaranteed peace and freedom for the Federal Republic of Germany. The military threat was present all that time and everybody knew that the ideological foe had strong forces stationed in the eastern part of Germany. In case of war we would have had to defend our own territory.

The political situation changed fundamentally in the years 1989 and 1990. Connected with the signs of internal disintegration within the Warsaw Pact, unification of Germany became possible. On October 3, 1990, the five states of the former German Democratic Republic were united with the Federal Republic of Germany. One aspect of unification was the integration of large parts of the East German Army into the Bundeswehr as well as the contract liability of the new Germany to continuously reduce the unified German armed forces to 370,000 soldiers at the end of the year 1994.

Figure 1

Personnel strength of the Bundeswehr

	October 1990	December 1994
Former Eastern Army	89,000	50,000
Bundeswehr in West Germany	432,000	320,000
Total	521,000	370,000

Even, the future German army is designed to remain a conscript force. This is the will of the political leadership and of the parties which support national goals. The supporters of conscription claim among other things that an all-volunteer force would be too heavy a burden for the German labor market and the national budget, and that the implementation of all-volunteer (regular) armed forces might lower the threshold for its use in military conflicts. Above all, the draft is regarded today – as opposed to the structures of earlier German armed forces – as a guarantor of the democratic spirit of the Bundeswehr and of its integration in society. Conscription is thus at the same time "a prominent instrument to further internal unity in Germany", because "the young people in East Germany are being conscripted and trained together with those in West Germany" (according to the radio station "Radioropa Info" on October 3, 1992, the second anniversary of German unification).

Longer-term planning provides for a marked increase in the relative portion of regular and temporary-career servicemen, however, it is planned that 155,000 conscripts shall serve in the Bundeswehr beyond 1995 (see *figure 2*).

Figure 2

Longer-term personnel planning (from 1995 on)

215,000	Regular/Temporary-Career Servicemen including 4,000 Reserve Spaces
155,000	Conscripts (42 %)

The basis of this planning is a trend analysis indicating that the quota of 155,000 draftees available for military service in an average age cohort of 370,000 seemed guaranteed permanently in spite of the increasing numbers of conscientious objectors and of young men who are exempted from the draft on account of exceptional regulations (e.g. three sons in one family or members of the police or of disaster control services; see *figure* 3).

Figure 3

Presumable quotas of conscripts in an age cohort
(Extrapolation for the 90s)

Average cohort strength (conscripts registered)		370,000	100 %
./.	physically unfit	66,000	18 %
	temporarily unfit	3,700	1 %
	not examined	3,700	1 %
Conscripts fit for military service		296,000	80 %
./.	Police service / Border guard service	8,000	3 %
	Disaster control services	27,000	7 %
	Other exemptions	23,000	6 %
./.	Conscientious objectors/ substitute service	52,000	14 %
Military service thereof		186,000	50 %
	Volunteers	14,000	4 %
	Conscripts	172,000	46 %

The validity of this long-term extrapolation has to be called in question even today. Sharply decreased numbers of applicants for enlisted service at the selection centers since January 1991 and unusually high portions of vacant posts at the quarterly induction dates for draftees substantiate considerable doubts that adequate personnel replacement in the armed forces is guaranteed. For example, as early as at the induction date July 1, 1992, almost every eighth position for draftees could not be refilled (see *Figure* 4).

Figure 4

Vacant posts for draftees

Induction date	July 1990	July 1991	July 1992
Required	52,047	51,797	50,100
Filled	48,573	46,624	44,075
Vacant	3,464	5,173	6,023
Vacant (%)	6.6	10.0	12.0

The numerous vacancies at the induction dates in the recent past were mainly caused by very high quotas of objections raised against induction orders. No less than 21,242 (32.5 %) of the total of 65,317 of the orders issued for the July 1992 induction date in order to meet the demand of 50,100 replacements had to be canceled for substantial reasons. Thus, only 44,075 posts could be filled (The impression that there is a lack of readiness to serve in the armed forces is increased by the large number of unfounded and therefore rejected objections).

Primary reasons for objections by conscripts

The conscripts had primarily the following reasons for objecting to induction:

- Worsened state of health (as compared with the pre-induction examination)
- Courses of study or vocational training had started in the meantime and discontinuance would be a hardship for the conscript prevented by law.
- The conscript is needed by his company (employer has to apply for it).

The high quotas of objections to induction orders are concomitant with a public discussion in Germany at present about the necessity and continuing meaningfulness of general conscription, which has been increasing since the turn of the years 1989 and 1990. What has contributed to this development is, on the one hand, the wide-spread feeling that, after the breakdown of the Soviet empire, Germany is no longer threatened militarily, on the other hand the terrifying media reports on military operations and effects of weapons in the Gulf war as well as in the crisis areas of Bosnia-Herzegovina, Azerbeijan, and Kurdistan. The uneasiness is increased by the realistic expectation that in future conflicts German conscripts will participate as UN soldiers. This is to say that, with increasing participation in United Nations peace-keeping missions, the German government will take into account the increased responsibility of Germany in the community of nations. The first visible steps in this direction are the German UN contingent in Cambodia, the participation of specialists from the German armed forces in the control of armament projects in Iraq, and participation in the emergency flights to Sarajevo and in the monitoring of the international embargo against Serbia. Among the public, however, these measures have not only met with acceptance and agreement; they have also aroused fundamental reservations, doubts and anxieties.

In marked contrast to the wide-spread expressions of insecurity and unwillingness to serve in the armed forces is the fact that the majority of conscripts have quite a positive attitude towards the armed forces' qualification testing. At our testing centers, only a very few conscripts refuse to cooperate or express their intent to object to military service.

In a representative opinion survey conducted in March 1992, the Psychology Service of the German armed forces questioned a total of 5,845 conscripts at their aptitude and placement test on their personal willingness to enlist for a longer term of service. The results of the study are not at all discouraging (see *Figures 5 and 6*).

Figure 5

Necessity of the Bundeswehr (%)

Urgently necessary	5
Necessary	37
Not so necessary	44
Unnecessary	14

Figure 6

Voluntary enlistment for a longer term (%)

Decided	5.1
Contemplating	17.3
Not at present	21.2
On no account	56.4

Another thing of particular interest was the frequency of answers to the questions concerning measures that could be helpful in improving the readiness to apply for a longer term, and how important conscripts regard being involved in determining the type, the place and the point of time of their military service (see *Figures 7 and 8)*.

Figure 7

Incentives for voluntary enlistment (%)

1. More money	39.0
2. Place of service according to wish	36.4
3. Military specialty according to wish	32.2
4. Higher payment after term	17.4
5. Occupational measures	16.1
6. Better promotion possibilities	15.0
etc.	

Figure 8

Desire to determine type and point of time of military service (%)

Yes, absolutely	92
Not so important	6
No, it does not matter	2

Altogether, the poll results show that the majority of conscripts are not opposed to military service at the time they take the aptitude and placement test, where they are normally 19 years old. They do, however, want to be involved in the decision about the place and the point of time of their call-up, and the type of military occupational specialty they are earmarked for. Now, the conscripts do not have the possibility to participate in their recruiting process in this way, in contrast to the volunteers. Significant reasons suggest that the high quotas of objections to induction orders may, among other things, be due to inadequate and outdated draft procedures.

Deficiencies in the draft procedure

The main deficiencies are:

- Only through the induction order do the conscripts get the information on whether, when, and where they are called up.
- The induction order is issued only a few weeks before the induction date; normally it thus comes unexpectedly and at an unsuitable time.
- Young men can be inducted up to the age of 25 (in exceptional cases up to the age of 32). For every second conscript the induction examination takes place more than two years prior to his actual induction.
- On fairness grounds, the conscripts are placed according to a computer-aided standard program. This program is not equipped yet to provide for the consideration of individual plans and interests.

Based on a psychological evaluation of these deficiencies, the armed forces' Psychology Service already some time ago derived the hypothesis that a considerable reduction in objections could be achieved by revising the procedure. A respective pilot study was conducted in spring 1992 by the psychology service at three regional recruiting offices. The study provided for conscripts to be individually counselled in an interview led by the psychologist and/or the responsible civil service official immediately after their aptitude and placement tests whenever this was feasible on organizational grounds. In this interview, the conscripts got an explanation of what military specialties in what places were suitable for them. In the frame of the military requirements, their individual wishes were taken into account in the best possible way for the decision. The place and point of time of their induction as well as the type of their military specialty were determined jointly. This procedure is much the same as the procedure for volunteers.

Because of the prevailing computerized rules and of the organizational constraints, a counselling interview of that type could only be led with a relatively small number of draftees.

Nevertheless, a comparison of the quotas of those not available for induction on July 1, 1992, because of exemptions showed that the "individually placed" conscripts did actually begin their military service considerably more frequently than conscripts who were placed according to conventional procedures (see *Figure 9*).

Figure 9

Posts of Conscripts not filled, Recruiting Offices Mannheim, Donaueschingen, and Heilbronn (July 1, 1992)

	Inductions	Exemptions	Percent
Conventional procedure	2,242	687	30.6
Individual placement	221	26	11.8
Total	2,463	713	28.9

The preliminary result of this pilot study is that the quota of objections can be more than halved simply by the early and personal involvement of the conscript in the placement decision. This is of so much importance that the Federal Office of Defense Administration has ordered a repeat of the analysis on a broader scale and under controlled conditions.

This study is at present being conducted in three major cities, München, Köln, and Hannover. Since problems in the replacement of conscripts have become more urgent than originally expected, a commission has meanwhile been established at the level of the Defense department and been tasked with developing a new concept for the draft procedures. In view of the experience gained in the pilot study in analyzing "individual placement" and in anticipating expected results of the control study just mentioned, the Psychology Service has submitted technical proposals for a new concept to the ministerial commission. These are above all:

- Early induction of conscripts, renunciation of the placement of older conscripts.
- Personal involvement of the conscripts in the decision about point of time and place of induction, and type of their military occupational specialty.
- Abandonment of automized placement, guarantee of individualized placement of conscripts where required.
- Intensification of contacts between military forces and induction official or psychologist at the regional recruiting office.
- Conduct of systematic evaluation studies into these measures.

Based on the personal rating of this speaker, the likelihood is rather high that these proposals will be accepted by the ministerial coordination group and be transformed into regulations of the armed forces administration. Specific aspects have meanwhile been realized. Whether or not the comprehensive catalogue of measures proposed by the Psychology Service is taken into account will surely depend on the results of the pilot study on individual placement. We very much hope that the trend of the first pilot study is confirmed and the number of objections drops considerably. In that case I think the defense administration would also agree to changes in the draft procedure, changes that have seemed advisable from the psychological point of view for a long time now. It would, among others, mean a better use of the Psychology Service's potential.

Equal Opportunity for East and West German Volunteers: An Unsolved Issue

Gerd W. Rodel
German Naval Volunteer Recruiting Center
Wilhelmshaven, Federal Republic of Germany

Introduction

People in East and West Germany were kept apart by the "Iron Curtain" for more than four decades. In these two German states two very different political systems developed. While a liberal democratic system was set up in the West, people in the East were oppressed by a totalitarian communist regime.

Life in these political systems has considerably changed the people in East and West Germany. The unification of the two German states in 1990 made apparent the differences between East and West regarding personality structures and confronted the people in East Germany in particular with a completely altered situation.

The political and social environment in the former GDR led to behavior patterns characterized by mistrust, introversion, insecurity and lack of decisiveness, strong norm orientation, avoidance of outstanding personal achievements, of responsibility and leadership tasks. The former GDR "rewarded" exactly those behavior patterns while repressing and "punishing" behavior not in line with communist ideology.

Precisely which differences may be expected from the East German candidates volunteering for service in the Navy who are confronted with the changed situation? How are candidates behaving in a selection process in which they have to present themselves successfully? Can the East Germans understand and accept the Western behavior patterns or "rules of the game"?

In answering these questions the individual developments in widely differing social and political systems or "part cultures" has to be taken into account – for up to that point everything from childhood onwards was regulated. The education and school systems in East Germany prescribed that all pupils had to have a certain basic knowledge (as for instance 2,000 words in orthography) "instilled" into them with no particular regard for the individual pupil. Marks given at school were at an inflationary high level, and in the final analysis depended among other things on the political attitude or adaptation of the pupil.

In accordance with these previous experiences, the Navy candidates from the East demonstrated the behavior patterns described. Day-to-day experience with this candidate population indicates that we should examine the development in various periods after unification.

In the *first phase* after unification, insecurity, skepticism, intimidation and distrust prevailed among the candidates. Many questions were put hesitantly and diffidently, for instance: "What do you intend with this question?", "I cannot commit myself!" or "Which answer do you expect to this question?" Justifications and self-accusations were uttered; self-confident demeanor and sureness were lacking.

The *second phase* – roughly nine months after unification – might be termed the "adaptation phase". The candidates endeavored to stay unnoticed, observed the behavior of the candidates from the West and copied their behavior.

It became apparent that the candidates from the East were very quick in learning certain social techniques of the West. In this phase, the Eastern candidates dared to clear up their unsureness by putting questions to the tester.

In the *third phase* – about 18 months after unification – the Eastern candidates often showed strange and sometimes shocking behavior. During CAT testing these candidates gave a bad impression, displaying disturbing, noisy and inconsiderate behavior.

It is not easy to assess or classify the behavior and outward appearance of Eastern candidates under conditions so completely unfamiliar to them. Although the Eastern candidates were treated with understanding and consideration by the raters in the selection process, it was often impossible to prognosticate positive development of the candidate in the Navy.

As prognoses of future development are normally based on past achievements explored by interviews and assessments, candidates from the East would clearly not have equal opportunities – particularly, if the raters cannot understand and evaluate the past adequately.

Moreover, Western selection procedures can scarcely do justice to these candidates because standardization and interpretation are not adapted to this new potential, so that capacity and capability of development are misinterpreted. This is shown very clearly in the case of candidates who have already proven themselves as petty officers first or second class in the GDR navy, but who can only present themselves with limited success in the selection process as candidates for the Federal German Navy.

With regard to the candidates' behavior in the CAT test situation, Eastern candidates in general were very reticent, timid and diffident, even discouraged. They needed considerably more time for the test instruction, some reacting in a hectic and restive manner, others with paralysing passivity; and they felt very oppressed and tense. Doubts and information gaps were not eradicated by putting questions to the tester.

Approach

In order to examine Eastern and Western candidates on a comparative basis, a random sample of 536 male Navy candidates was assembled, taking care to balance the two groups approximately with regard to age, schooling and vocational level.

The data were collected within an Assessment Center which every candidate has to pass through. Here the results of the achievement tests (corresponding roughly to the CAT ASVAB), of a physical fitness test as well as ratings concerning personality characteristics from the interview were compared.

Results

Differences between candidates from East and West Germany

Under the basic conditions already described the differences in the selection of candidates from East and West Germany turn out largely as expected:

In nearly all subtests significant discrepancies appeared in the test results when comparing the Eastern and Western groups. That is, the Eastern candidates clearly solved fewer

items in the general achievement tests under pressure. The differences are less significant in the more practice-oriented tests, as for instance mechanical tests or reaction tests.

Furthermore, the Eastern candidates turned out to produce rather fewer incorrect solutions in the tests.

Eastern candidates needed considerably more time for the test instructions, and they also took longer to deal with the individual test items. These differences did not apply to the above-mentioned practice-oriented tests, however.

In the physical fitness test, on the other hand, the Eastern candidates outdid the group from the West.

The differences in the rating of traits were less significant. The Eastern candidates were only given lower scores in the traits "motivation to perform", "intelligence", "comprehension", "articulateness" and "physical fitness".

The overall military qualification, which is derived from the total achievement and the evaluations of the Assessment Center, did not show significant differences between Eastern and Western candidates over the last two years.

Changes over the past two years

These results, which have been compiled over the whole period since unification, conceal changes which have emerged in the performance behavior of the Eastern candidates over the past two years. When the results are considered over time, considerable changes do become apparent. In order to give a general idea, the results have been set out on a quarterly basis:

Test results. At the end of 1990 the Eastern candidates started out with considerable deficits in correct solutions in the tests. They behaved hesitantly and indecisively in solving their tasks, often asked questions or behaved passively and were ultimately unable to meet the time limit. After only half a year the results grew noticeably better and kept going up continuously, although up to the present they are still significantly below the level of Western candidates. West German candidates remain constant at an average of 21 correct test solutions.

Time for Instruction. At the beginning of the testing the candidates are given a general standardized instruction via screen and headphones. The minutes which candidates took on an average to work their way through this instruction were counted. There was no time limit on this. At first the candidates from the East took considerably longer for this familiarization with the test. On average they needed three more minutes. This difference has by now been reduced to half a minute. There are also far fewer questions being put to the tester so that the time for instruction is not lengthened under this aspect either. A similar trend is to be found when we compare the instruction times for the various subtests.

Incorrect solutions. As for the trend of incorrect test, Eastern candidates started out in 1990/91 with an average of 11 errors measured in all tests. This score dropped below the score of the West Germans about one year later but is now rising above 11 errors again, while the candidates from the West remain fairly constantly at a level of 9 errors. Recently Eastern candidates have been approaching their tasks in a much more superficial and indifferent manner, showing little serious endeavor in some cases.

Military Qualification. In considering the military qualification over the last two years, some differences only – as mentioned above – were found between Eastern and Western

candidates. Taking a more detailed look at the individual quarter years, differences between East and West become apparent.

At first the Eastern candidates behaved suitably and better adjusted and caused less disturbance. Probably for that reason they were given a more positive military qualification for the Navy. Since October last year the opposite has been shown by the Eastern candidates, who are drawing attention to themselves by a conspicuous and forward behavior. Because of these peculiarities of behavior they are then given a less favorable military qualification prognosis.

The West German candidates, on the other hand, remain at a constant level except for seasonal variations due to the graduation of secondary-school students in the summer and of vocational trainees in the fall.

Evaluation of results

If the results are to be interpreted adequately, it has to be taken into account that the Western candidates have a certain amount of experience and a conception of what is comprised in an application, how to "sell" themselves, to present themselves as socially desirable and to promote themselves. They also have some idea of the demands to be met in the new sphere of activity. Thus a certain self-selection takes place among the Western candidates from the outset.

The candidates from the East have not the faintest idea of what is in store for them in a selection process, however. Some of them express completely exaggerated expectations, imagining that the Navy is bound to employ them, give them at least a driver's licence and vocational training, or even the opportunity to go to university. These after all, were the things the GDR Navy offered if one volunteered for it as a good GDR citizen who toed the line.

In retrospect they cannot comprehend, why the West German Navy rejects them now. A self-critical analysis in the sense of self-selection does not take place. Also, it happens only very rarely that they apply for several jobs.

Methodic distortions cannot be excluded in this, either owing to the self-selection process mentioned above with regard to the personality characteristics of the population having remained in the GDR for one thing, and in connection with possible differences in the motivation behind joining the Navy between East and West for another. The results might also be distorted because of the change of system with its consequent loss of values, difficulties in orientation and loss of motivation. The various causes are probably mixed together. It is practically impossible to examine them separately, however.

The lesser achievements of the Eastern candidates may also be due to the fact that this population of candidates has little familiarity with and experience of tests, particularly experience in coping with CAT tests.

If up to now we have been treating test results as objective data and system-independent facts, we have to realize after all that these data, too, can be influenced to a considerable extent by experience and situational conditions – even in a computer-aided environment.

The causes for the differences in the personality area between Eastern and Western candidates presumably depend mostly on experience and learning. It may be assumed that the deficits in competence among East Germans are due not so much to personal characteristics affecting their achievements but rather to moulding by different systems. Yet, if this is true, these learning processes must be reversible, i.e. the people in East and

West will reach to a common level in the foreseeable future. The same can be seen from the results of other analyses in the civilian sphere. On the other hand, there are processes at present in the East which are to be interpreted rather as a tendency towards distinctiveness and self-sufficiency, in direct contrast to "unification or adaptation" in its true sense.

The conspicuously negative behavior of Eastern candidates during recent selection procedures obviously corresponds to the markedly radical and atypical behavior of young people in the East, which takes the form of riots and outrages against foreigners. Recent events also show that changes are occurring more quickly than expected but taking a negative turn.

Conclusions

What may be done to assess candidates from the East more appropriately and to provide them with better starting conditions for their entry into the Navy?

- Evaluators ought to be trained in dealing with candidates from the East and given a well-founded basic knowledge about conditions in the former GDR so that they learn to understand better the facts and – on that basis – the behavior of the people.
- Fact-finding visits to the East ought to be made in order to bring about an understanding of the situation of upheaval, the high unemployment rate and disorientation in the East.
- Candidates from the East could be informed more thoroughly and in a manner more readily understandable to them about what is in store for them in the Navy before the actual interview.
- The raters ought to be trained to enable them to judge candidates from the East not only according to Western standards but also by taking into account their background in the former GDR. Moreover they have to learn to assess correctly the biographical data, curricula vitae, diplomas and documents.
- The raters have to reach an understanding of the situation of the candidates who have been put in the position of the "losers" by the quick collapse of the GDR system and feel derogated.
- In assessing a candidate, one ought to take into account not only the past development of every single candidate, but also the present developments in a difficult phase of upheaval and disorientation.
- Separate test standards might be set up for Eastern candidates, which would have to be continuously adapted though. In this, aspects of validity and efficiency in military service should not be disregarded, even if Eastern candidates are judged on a different development basis than Western candidates.

By these measures, greater fairness and equal opportunity can be achieved for the candidates from the East.

Aptitude and Motivation of Officer Applicants Selected and Rejected

Albert H. Melter
Federal Armed Forces Central Personnel Office
Cologne, Germany

Word has repeatedly come from the services that the junior officers in the Federal Armed Forces do not have the motivation and attitude it takes to take up a military profession. The two-pronged procedure, comprising an aptitude test and placement in an officer career, is said to produce a situation where the "wrong" officer candidates are enlisted. The officer candidates are not so much being criticized for a lack of intelligence. The main accusation made is that they do not have an adequate approach to military service. It is the selection procedure that is under attack, not the quality of the applicants, not the quality of training and education in the forces, not the quality of motivation by superiors.

One argument supporting the alternative hypothesis that not enough of the "right type" of officer applicants are enrolling is that the battalions have considerable numbers of reserve officers whom their commanders speak well of. For instance, paratroop reserve officer candidates are said to have the kind of attitude towards their military function that the forces would particularly like to see among active officer candidates. Unfortunately, however, many of these ask to be relieved just a few days into the ranger training course because they have neither the staying power nor interest to see it through; this is at least the way those on line appointments account for a phenomenon that has been observed for years now in officer education and training.

Junior officers are selected to specifications issued by the military staffs and the Personnel Directorate of the Federal Ministry of Defence. The mission performed by the Federal Armed Forces Central Personnel Office is to implement the aptitude test and career placement procedure. It is also responsible for examining whether or not the lack of motivation and wrong frame of mind that are complained about can already be detected while applicants are still at the aptitude test stage. The purpose of this analysis is to increase the level of knowledge reached in the spring of 1992 when a poll was conducted among conscripts and volunteer applicants regarding the attitudes and opinions they have affecting their willingness to sign up as volunteers.

Willingness to Enlist as Volunteers

On account of the Gulf War, the structural changes within the armed forces and their corresponding media coverage, and uncertainty about the Federal Armed Forces' new role, applicant figures in the year 1991 were affected by a host of external influences; so to get a better idea of the situation, it would be wiser to compare the figures for 1990 and 1992. It is evident that while applicant figures continue to decline, the proportion of applicants from the new Länder exceeds the proportion of the population that lives in these Länder. The main reasons for this drop in applicants were cited in the poll conducted among conscripts during the 1992 pre-induction examinations. They are, arranged in the order of the frequency with which they were quoted and from the point of view of the conscripts themselves (Federal Ministry of Defence, 1992): industry and commerce offer better career prospects; the lack of certainty about being assigned anything but a local posting; fear of "too many constraints" in

military service; too little scope to improve earnings; the risk of being employed in military activities outside the Federal Republic of Germany.

With the number of applicants far exceeding the number of posts provided for in the 1992 budget, which has declined considerably in comparison with the previous year, the likelihood is that the personnel requirements for 1992 can be met with personnel of a good standard, despite the problem arising from the fact that there are less young men coming forward to enlist.

The Federal Armed Forces Academy for Information and Communication in 1991 summarized opinion poll data on security and defence policy in Germany (Hoffmann, 1992). The information compiled revealed a widespread attitude towards the Federal Armed Forces: "We" need the Federal Armed Forces - "I" do not, though! The Federal Armed Forces are considered a useful institution while they concern the state. Whether or not the individual has a personal need for the Federal Armed Forces is debatable. Young men, however, are willing not only to do basic military service, but also to enlist as volunteers when they are affected by the armed forces personally in a positive sense. The 1990 SINUS survey of young people showed that 41 % of the 16 to 24-year old age-group certainly wanted to do basic military service, 10 % wanted to enlist as volunteers, 7 % were not sure what kind of service of the two they were going to choose and 2 % had already volunteered. This means that altogether, 59 % had decided in favour of the Federal Armed Forces (Hoffmann, 1992).

The young men and women who submit applications to become volunteer officers in the Federal Armed Forces are tested by officers, qualified psychologists and doctors at the Federal Armed Forces Central Personnel Office in Cologne to determine their aptitude and motivation for the training involved and a longer-term enlistment in a military command function. This is where applicants are assessed and selected for all service assignments and all subjects of study offered by the Federal Armed Forces' two universities. Placement in a particular career, in most cases a combination of assignment and subject of study, is always conditional upon an applicant's general aptitude. Applicants considered to be unsuitable for officer training are counselled and assisted in obtaining information on other training schemes offered by the services, and, if they are interested, in applying for them.

Samples

The aptitude and motivation of officer applicants selected and rejected have been studied for several complete applicant age-groups. The years selected were 1988 to 1991, in order to inlcude both the year preceding the political changes in Germany and Europe, the years of the turnaround in and unification with Eastern Germany, the Gulf War and the onset of the conflicts in the southeast of Europe in the analysis. For each of the four applicant age-groups, the data ascertained was the information of relevance on all the applicants came out of the aptitude test with the ratings of "well-suited", "suited", "suitable with limitations" or "unsuited". Applicants who by choice did not see the aptitude test through to the end were not considered in the sample.

The applicants selected were officer applicants suited for careers as line and medical officers, regardless of whether they were indeed placed in such a career or whether they were not enlisted on requirement grounds. The applicants rejected were those who failed to pass the aptitude test.

Figure 1

Sample Size

Age-group	Selected	Rejected
1988	2,234	3,823
1989	2,515	3,795
1990	2,455	3,989
1991	1,900	2,832

As part of a small sample made of 397 officer applicants in the spring of 1992, an opinion poll was conducted similar to that among the conscripts (MOD GE, 1992) to examine in a pilot study the attitudes the applicants had towards the role of the Federal Armed Forces, possible future tasks (for example, for the United Nations), and measures aimed at enhancing the willingness among young people to join the services. It must be borne in mind that this sample was a small one and with regard to the applicants enlisted in 1992 maybe not representative. Consideration must also be taken of the fact that the applicants were in a selection situation and that some 40 % of the group were unsuited.

Method

The data was classified in accordance with the variables of applicants selected and rejected and those applying for enlistment between 1988 and 1991 and with the careers of line officer (LO) and medical officer (MO). The categories were described with frequency distributions in percentages and compared using the dependent variables of general aptitude, aptitude for subject of study, concepts of the profession, eagerness to learn and latent vitality, ability to tolerate mental strain.These are some of the points for which every applicant is examined when his aptitude is being assessed and rated.

There are four grades to the general aptitude variable: An applicant is considered well-suited (A) if he is rated satisfactory (4 on a 7-point scale) or better on all 11 characteristics used to dete: mine his aptitude. An applicant is considered suited (M) if he is rated adequate (5 on the 7-point scale) or better on all characteristics. An applicant is considered suitable with limitations (X) if he is rated below-average (6) or inadequate (7) on one or two characteristics. Finally, an applicant is considered unsuited, either after he has completed all the tests in two days (Z) or after the first day's tests (Y) when he is rated below-average or inadequate on several characteristics. There are three grades to the subject of study recommendation variable: recommended; recommended with limitations, successful completion of studies doubtful.

The concepts of the profession are defined as the willingness to perform the tasks and duties of an officer and as the degree to which applicants have concepts of military necessity. The variable in this survey is rated as a measure of the motivation an applicant has to volunteer for the career of an officer. The other motivation variable is defined as the aptitude characteristic "learning and achievement motivation" based on the ability to set a fair target for something, diligence, endurance and aspiration for success. The third variable concerning aspects of motivation is the aptitude characteristic "stress resistance". This is defined as the ability to remain willing to do something or retain self-control, a high standard of performance and mental balance when under mental and social or situation-related strain, though notably as the ability to come to terms with failures and disappointments. These aptitude requirements are but part of what is asked of officer applicants. They are observed, assessed and

rated on a scale from very good (1) to inadequate (7) by a selection commission made up of a qualified psychologist and two officers.

Results

In the four years from 1988 to 1991, the number of officer candidates required for the career of line officer had to be made up on an average of 41 % by applicants who were rated suitable with limitations. A wide variety of reasons prevented the Federal Armed Forces from enlisting a quarter of the applicants who were well-suited, a third of those who were suited and more than half of those who were suitable with limitations to meet requirements. The main reason was that there was no need for the combination of assignment, subject of study and term of enlistment the applicant wanted and he was not prepared to consider any alternatives.

Figure 2

Percentage of unsatisfactories in the aptitude requirements of importance for motivation

Age-group	1988	1989	1990	1991
Concepts of profession*	56	61	56	57
Learning and achievement motivation **	49	54	43	41
Stress resistance	54	59	43	44

* Until 1989 described as "sense of responsibility"
** Until 1989 described as "dedication, drive"

In terms of the aptitude requirements important for determining whether or not an applicant is later enlisted for service as an officer, an average of 51 % of the officer candidates taken on for careers as line officers were rated unsatisfactory while they were applicants, that is to say, lower than 5 on the 9-point scale used up to 1989 and lower than 4 on the 7-point scale used since 1990. Of those considered relatively the "best", those found to be suited, there were many who by no means were rated good.

Regardless of whether they could be enlisted or not, the percentages among the applicants selected of officer applicants who were suitable with limitations rose from year to year, while the percentages for those applicants who were suited dropped. Even among the selected applicants finally enlisted, the percentage of those who were suitable with limitations rose from 1988 to 1990 (1988: 36.3 % / 1989: 40.9 % / 1990: 43.6 %). In 1991 it dropped to 40.1 % again. These trends could not be noted among the medical officer applicants selected. There was a drop in the percentage of suited candidates among the officer applicants rejected. From 1988 to 1990, there was also a rise in the percentage of applicants selected whose prospects of successfully completing a course of study at the Federal Armed Forces' universities were considered limited or doubtful (1988: 55.6 % / 1989: 58.2 % / 1990: 60.6 %). In 1991 it dropped to 55.3 % again.

Figure 3

Trend towards an increase in limitations in aptitude among officer applicants altogether (in %)

Aptitude		1988		1989		1990		1991	
		SEL	REJ	SEL	REJ	SEL	REJ	SEL	REJ
Well-suited	OA	5.6	1.4	4.4	1.5	6.5	1.4	8.0	1.8
	Med OA	19.7	3.0	37.4	3.0	56.5	2.9	28.7	2.5
Suited	OA	55.1	5.9	50.7	16.5	46.3	11.8	43.6	8.8
	Med OA	73.9	23.4	56.4	37.9	43.5	33.4	62.9	25.8
Suitable with limitations	OA	38.7	28.7	43.3	21.2	46.6	23.0	46.9	21.5
	Med OA	6.4	43.6	6.1	38.2	0.0	38.8	8.4	44.4
Unsuited	OA	0.7	53.9	1.6	60.7	0.7	63.8	1.5	67.9
	Med OA	0.0	30.0	0.0	20.9	0.0	24.9	0.0	27.2

The proportion of applicants whose concepts of the profession were limited (rating of 6 on the 7-point scale or 7 on the 9-point scale) or inadequate (rating of 7 on the 7-point scale or 8 and 9 on the 9-point scale) remained the same throughout the four age-groups – at around 25 % –, the only exception being among the medical officer applicants, where there was a sharp drop . Compared with 1990, there was a marked increase in 1991, the year of the Gulf War, in the percentage of applicants whose concepts of the profession were slightly limited (rating of 5 on the 7-point scale or 6 on the 9-point scale). Trends of this kind were not noted for the aptitude characteristic "learning and achievement motivation". As for the aptitude characteristic "stress resistance", there has been a steady fall since 1989 in the percentage of applicants selected whose ability in this field was limited or inadequate.

The opinion poll conducted in March 1992 among 397 officer applicants revealed that the majority had a positive attitude towards tasks performed by the armed forces, even to tasks the Bundeswehr may assume in the future. For example, no more than 2 % rejected the idea of the Federal Armed Forces taking part in UN peacekeeping operations, a mere 9 % rejected their involvement in UN force armed action. 60 % rejected the view that the Federal Armed Forces should be employed solely in the defence of our country. The applicants also expressed firm opinions on the pros and contras of volunteering for service in the Bundeswehr, for instance, on the reasons given by people of their same age against doing so. When asked why people of their same age did not apply for a career in the services, 61 % voiced the opinion that there was more future in industry and commerce, and 42 % said that there was little public prestige in being a soldier. Other excuses given for people of their age were the possibilities of being stationed far away from home (67 %) and of being employed in military activities that exceeded the limits set by our country's Basic Law (52 %); in addition, that military service did not appeal to them (40 %), that service personnel had to accept too many restraints or obligations (48 %), and also private, family reasons (49 %). When asked what measures should be taken to recruit suitable candidates, the interviewees expressed 14 preferences: 1. Preferable posting and 2. assignment. 3. More money. 4. Better chances of promotion. 5. Enlistment after a probationary period. 6. Support in gaining qualifications for civilian careers. 7. More courses of study to choose from. 8. Large enlistment bonus. 9. Eventful service. 10. Greater promise

of becoming a regular. 11. Larger severance payment. 12. Use of advanced technology. 13. Comradeship. 14. Better provision for physical education. What is evident here is that the dominating attitudes in society, namely, self-determination and materialism, outweigh other motives.

Discussion

The results show that it is at least questionable to ask too much of the future officer candidate with regard to his concept of a military career or profound knowledge of the subject. Instead, for an officer candidate to become an officer in a high-tech conscript army in a democratic and pluralistic state, he must undergo training and education in line assignments and at the Federal Armed Forces' various schools. If in the course of training it should turn out that the officer candidate does not come up to scratch, the procedure to discharge him on the grounds of inaptitude can be initiated (Melter, 1991). From 1988 to 1991, 4 % of the officer candidates enlisted were later discharged on medical and social grounds and 5 % for personal or professional inaptitude.

With the general opinion at home, at school and at work being that the Federal Armed Forces are no different to any other employer, people with an intrinsic motivation to perform military functions are few and far between. Virtually all young people in our society, even officer applicants, expect employers to provide them job security, good professional or vocational training, with a minimum demand for mobility. The Federal Armed Forces have responded to this by adopting the principle of local postings for conscripts and regional transfers for volunteers.

If today's junior officers fail to satisfactorily meet expectations in the ranger training course or other comparable tests, this is only marginally due to the fact that too many mistakes are made in the selection procedure on account of the particular emphasis placed on objective criteria. For no matter what other method we used, we would still have to work with young men and women who are brought up and live in a world that is not exactly conducive to the development of certain characteristics an officer needs and who are attracted by an advertising régime that aims at precisely those attitudes and values they have which are now the cause for complaint. We must base our selections on the applicants we get.

References

Federal Ministry of Defence, P II 4 (1992). *Meinungsbefragung von Wehrpflichtigen und Freiwilligen vor Eintritt in die Bundeswehr (Poll among Conscripts and Volunteers before they joined the Federal Armed Forces).* Unpublished and interim working report dated 20 May 1992. Bonn.

Hoffmann, H.-V. (1992). *Demoskopisches Meinungsbild in Deutschland zur Sicherheits- und Verteidigungspolitik 1991 (Opinion poll data on security and defence policy in Germany in 1991).* Waldbröl: Federal Armed Forces Academy for Information and Communication.

Melter, A. H. (1991). Discharge During Officer Training and Consequences for the Selection Strategy. In *Proceedings of the 33rd Annual Conference of the Military Testing Association* (pp. 610-615). San Antonio.

Computer-Assisted Programs in the Training of Leadership Behavior

Eckhard W. Bucher
Non-Commissioned Officer School
Münster, Germany

Introduction

The demands on a young non-commissioned officer are wide-ranging and have constantly increased in the last few years:

- the non-commissioned officer leads his soldiers and his weapon system in peace as well as in war times and thus under extreme strains,
- as a military instructor he bears a share of responsibility for the training level of his group,
- and as a military instructor he must be willing to function as an example and thus be able to illustrate the sense of military training.

Frequently the public regards his achievements as those of the army as a whole. Therefore, it is important that the young non-commissioned officer should have a high level of training and education. Especially the constantly increasing demands on his skills to lead other people require qualified training.

The German Army Non-Commissioned Officer Schools

So far four German Army Non-Commissioned Officer Schools have been installed to train young non-commissioned officers and to prepare them for their tasks in the army. This shows the importance attached by the Federal Armed Forces to a modern way of leading soldiers.

I am going to describe some aspects of the "Training Course for Sergeants Part I". This course aims at enabling the young non-commissioned officer who is often hardly older than the soldiers under his command to effectively fulfil his tasks as a military leader by educating and training his soldiers according to modern standards of adult education.

As a rule, the course takes place at one of the four German Army Non-Commissioned Officer Schools after the young men have been soldiers for about 15 months. It lasts 8 weeks and is based on the knowledge and experience the young non-commissioned officers have acquired as military leaders during their service in the army.

One aspect the course focuses on is a seminar called "Modern Leadership". The center of this seminar is the training of adequate modes of communication and behavior as an indispensable prerequisite for a modern way of leading people. The student is expected to recognize that effective leadership behavior is not a mysterious art which has been given to you by God but that this behavior can be learned and trained.

Developments in the Field of "Computer-based Training" (CBT)

Everybody knows that in our modern industrial and educational society the computer plays an important part and that it is very hard to imagine life without it. As a consequence of the explosive development of software and hardware supply, the methods of "Computer-Assisted Instruction" (CAI) respectively "Computer-based Training" (CBT) become more and more important.

The use of the computer in training and education aims at teaching at least part of the learning matter with the help of computers: on the one hand by relieving the staff, on the other hand by imparting various learning matters in an illustrative way. This is important because our knowledge becomes outdated faster and faster and the technical and social changes and developments which nowadays happen more and more quickly entail an increasing demand for further training and learning.

Computer-based Training (CBT) at the German Army NCO Schools

As in many civilian enterprises, the computer is being used as a training means also at the German Army NCO Schools. Recently, so-called "interactive computer-assisted training means" have been introduced for teaching soldiers. This is a computer-assisted system of information which enables a student to learn while having a dialog with a teaching system. This system is a learning program developed according to the principles of learning psychology, and it is presented to the student with the help of a computer. At present four of such learning programs are in use in the fields of "Training Method" and "Leadership Behavior".

In the course of the seminar "Modern Leadership Behavior" the students are expected to develop modes of communication and behavior necessary for a modern way of leading soldiers. In this field the "dialog between "superior" and "subordinate" is not only an important means of communication but also a very effective means of leading other soldiers. The students are to gain this knowledge by thoroughly going through the learning program entitled "Soldier Renner is always ill on Mondays". As all four learning programs are similarly structured, I am going to describe and explain this learning program in detail.

"Soldier Renner strikingly often rings in sick on Mondays. Last Monday he did not appear for his military training after he had returned from a medical check-up without having been found ill. Sergeant Schilling is Soldier Renner's superior and group leader. Schilling has only recently been appointed. Now he is to react to Soldier Renner's behavior."

At first the student is expected to imagine Sergeant Schilling's part who acts in the student's place. Then the student is offered various reactions in a menu by the computer. Here are some examples:

- showing understanding and waiting,
- reporting the case to the platoon leader,
- asking Renner's comrades,
- having a dialog with Soldier Renner etc.

The student can see these possible reactions in video clips. Having decided for one of these reactions, he will see the effects of his decision in another clip.

With the help of this first menu the student is expected to recognize that he must be fully aware of his surroundings. "Perceiving consciously" Renner's behavior means: Realizing that something is wrong! As soon as the student is aware of this , the possible reaction "showing understanding and waiting" is the wrong decision. The student who chooses this possibility nevertheless is confronted with the fact that Soldier Renner will not change his behavior at all.

However, if the student has fully perceived "the Renner problem", his first idea will be: I must do something!

Now the student can choose among two possibilities for his further actions:

1. The first one consists in his attempt to pass the buck to his superior by reporting the "Renner problem" to him. In this case the Sergeant does not make use of his function as a superior and group leader and does not meet the responsibility given to him. This is why the platoon leader reacts sharply and does not give any advice or help after Sergeant Schilling has informed him of Soldier Renner's behavior. He orders Sergeant Schilling to try to handle the "Renner problem" first on his own, using all the possibilities he has as a group leader.

2. In case the student decides to choose the second way, he will aim at solving the "Renner problem". His activities will focus on getting information on the reasons for Soldier Renner's behavior. Here the learning program offers four choices:
   - phoning the medical staff,
   - asking Renner's comrades,
   - conversation with another sergeant who knew Renner formerly,
   - dialog with Soldier Renner.

The first three ways of action lead Sergeant Schilling into a dead end. The video clips show this clearly in the reactions of the persons concerned. On the contrary, this choice involves the danger that Sergeant Schilling may get distorted information, subjective statements and prejudices. It is almost impossible to distinguish between subjective opinions and hard facts when questioning the persons involved. In addition, this way is time-consuming.

The fourth possibility teaches the student that it is often better to get one's information directly, i.e. Sergeant Schilling has to speak to Soldier Renner. As soon as the student has made up his mind to speak to Soldier Renner he has to decide in the following steps of the program how to start and to continue the dialog etc.

As can easily be seen, the learning program is widely branched and is directed by the decisions of the students. The learning program gets its structure by means of the menus in which the various choices of acting are listed. The student can see these possibilities in video clips before he makes his decision for one of them. After each decision he learns – with the help of a video clip – how the persons involved react to his behavior. Now it is his turn again to react to this reaction, and so on.

Whenever the student realizes that his choice is not successful, respectively leads to a dead end, he has the possibility of altering his decision. It is not decisive to finish the learning program in the shortest time possible and to choose the ideal solution. Rather, it is important that the student learn to know the various possibilities of action with all their advantages and disadvantages and to evaluate them critically.

The learning program does not offer any solution of the "Renner problem". On the contrary, the learning program focuses on how a dialog as a means of leading others is used. The students are to learn that a dialog correctly carried out will result in confidence and understanding among the participants of this dialog. Thus, the learning program ends with the fact that Soldier Renner gains confidence in his superior and is willing to talk about the reasons for his behavior.

Some sort of testing takes place as the student is ordered again and again to analyse behavior and to act adequately. Only if he is constantly successfully doing so, will he proceed in the learning program. In this way he himself controls the success of the behavior he has chosen without having the impression of being tested.

While the student goes through the learning program, all his data are stored on a floppy disk under an individual password. These data can be evaluated by the teacher and discussed with all the students immediately after the presentation of the program. In this way the student gets a feedback of how the other students went through the program. Thus, the students are enabled to discuss the advantages and disadvantages of the various possibilities of behavior.

Evaluation of the Learning Programs

Since the "Renner Learning Program" as a means of training has been in use for a short time only, no systematic investigations are at hand yet. That is why I have to confine myself to some provisional remarks. These are based on qualitative comments of some hundred students who worked through this program.

1. The students appreciate the learning programs for the field of "leadership behavior" and the acceptance of the programs is rather good.
2. All the students value highly that the whole learning process is informative and entertaining as well and that the student can to a large extent choose the speed of his work and his own way. Besides, all students appreciate being able to make their decisions without time pressure, which happens rarely during practical training. In this way well-considered acting can be trained.
3. After having been worked through, the learning program must be intensively dealt with. The students want to discuss the advantages and disadvantages of their choices in detail. If this is not done, the student feels "spoon-fed" and restricted by given choices in his freedom to make his own decision.
4. Much depends on how "realistic" the students judge the learning programs: the acceptance of the learning programs, moreover, that of the whole "Computer-based Training (CBT) increases with the degree of reality attributed to the learning programs. A learning program is considered the more realistic"
   - the more the programs are identical with everyday situations in military life and applicable to everyday routine,
   - and the more the student in the programs realizes experience he was confronted with in the army himself.
5. From the beginning to the end the different roles in the individual video clips should be played by the same actors so that the student can better realize the models of behavior;

besides, there should only be a few actors in the video clips so that an identification with the main characters is possible more easily.

6. The military ranks ("grades") of the persons acting should be similar to those of the students.

Prospect

Summing up one can say that so far the results of the use of computer-assisted learning programs in the training of effective "communication" respectively "leadership behavior" at the Non-Commissioned Officer School in Münster are very encouraging. Up to now four learning programs have been in use at the NCO Schools in the fields of "training method" and "leadership behavior". In the future other computer-assisted learning programs are planned to be used in the fields of "Politics", "Group Dynamics" and "Communication".

BUILDING A JOINT-SERVICE CLASSIFICATION RESEARCH ROADMAP: CURRENT CLASSIFICATION PROCEDURES

Teresa L. Russell and Deirdre J. Knapp
Human Resources Research Organization
John P. Campbell
University of Minnesota and Human Resources Research Organization

This paper provides an overview of military enlisted personnel selection and classification (or assignment to jobs). Information for the paper was gathered in interviews with military selection and classification research experts and through publications documenting military selection and classification procedures. Between January and April 1992, we interviewed 43 selection and classification experts from the Armstrong Laboratory, the Army Research Institute (ARI), the Navy Personnel Research and Development Center (NPRDC), the Center for Naval Analyses (CNA), the Military Accession Policy Working Group (MAPWG), the Defense Manpower Data Center (DMDC), and the Office of the Assistant Secretary of Defense for Force Management and Personnel (OASD-FM&P). The sample consisted of the professional, scientific, and management personnel from these organizations who are most concerned with selection and classification issues.

The interviews had several goals: (1) to brief the participants on the Roadmap project, (2) to develop a list of objectives for classification research over the next ten to 15 years, (3) to learn more about the context (mission, applicant pool, occupational structure) in which selection and classification decisions are made currently and how that context may change in the next ten to 15 years, (4) to learn about steps currently being taken toward the accomplishment of the research objectives, and (5) to begin gathering specific information for subsequent Roadmap tasks focusing on predictors, criterion variables, job analysis procedures, and statistical methodologies.

We organized the information from the interviews into three areas: (1) the current selection and assignment procedures used by each of the Services; (2) the military classification environment and factors that affect the assignment systems; and (3) our discussions with interviewees yielded a list of 25 selection and classification research objectives. We asked interview participants to make judgments about the importance of those objectives. This paper reviews 1 and 2 above. John Campbell will begin his talk with some data regarding experts' judgments about the objectives.

Selection and Classification Procedures

Each Service develops and applies its own selection and classification procedures. Selection procedures--pre-screening, aptitude testing, and medical examination--are very similar across the four Services. The major distinctions among the systems occur in assignment to jobs, where each Service applies its own criteria for making a Person Job Match (PJM).

The Recruiter's Role. Recruiters play an important role in the identification, attraction, and selection of qualified recruits. Recruiters also administer an abbreviated version of the Armed Forces Qualifying Test (AFQT) to pre-screen prospects on verbal and quantitative aptitude, the Enlistment Screening Test (EST). It is a paper-and-pencil test used by all the Services. When computerized testing is possible, Army recruiters use the Computerized Adaptive Screening Test (CAST) instead of EST to pre-screen applicants.

Mobile Examining Team (MET) Sites and Military Entrance Processing Stations (MEPSs). The next step in the enlistment process is the administration of the Armed Services Vocational Aptitude Battery (ASVAB). Recruiters either personally transport or send prospects who appear to be suitable for service to a Mobile Examining Team (MET) site or Military Entrance Processing Station (MEPS). MET sites are small ASVAB testing centers that are distributed across the United States; there are about 900 MET sites. MEPSs are larger stations where full-scale enlistment processing is accomplished (e.g., medical examination, counseling, HIV testing). MET sites are more accessible (and thus less costly) than MEPSs.

At this point, the Army, Navy, and Marine Corps screen applicants on AFQT, a measure of verbal and mathematical ability. Scores on AFQT are reported within five broad AFQT categories based on percentile score ranges. The Army's operational cut score is at the 31st percentile on AFQT, for applicants who are, or will soon be, high school graduates. The Marine Corps screens at the 31st percentile on AFQT and requires that applicants meet a minimum requirement on a General Technical (GT) composite. Similarly, it is Navy policy to screen applicants at the 31st percentile. The Air Force uses a somewhat different strategy. Applicants are screened on AFQT, but the cut score is low enough (31) that most applicants qualify. The additional, more stringent screen, involves summarizing applicants' ASVAB scores according to four composites: Mechanical (M), Administrative (A), General (G), and Electronic (E). The sum of the M, A, G, and E percentile scores must currently be 185 or higher and the G composite percentile score must be at least 45 for entry into the Air Force.

Ultimately, applicants are sent to a MEPS where they undergo physical and psychiatric evaluation. Doctor's assessments of applicants are assembled into a code called PULHES. In addition to the PULHES evaluation, applicants are tested for HIV antibodies at the MEPS; HIV positive applicants are not admitted. Until recently, both the Army and the Air Force administered strength tests at the MEPSs. The Army has terminated use of its test, the Military Entrance Physical Strength Capability Test (MEPSCAT), which measured the amount of weight that an applicant could lift. Currently, the Air Force is the only Service using a strength measure, called the X factor, which is comparable to MEPSCAT.

Occupational/Job Assignment. All Services assign applicants to either an occupational area or a specific job at the MEPS. Although the process differs somewhat across the Services, generally a career counselor, or classifier, reviews the recruit's aptitude scores, medical history, and educational records. The counselor uses a computer

system to obtain a list of current and future technical school vacancies and specialties, in order of Service priority, that match the applicant's records. Applicants and counselors discuss the job options, and the applicant makes the final decision about enlistment (Camara & Laurence, 1987).

Aptitude scores are an important component in each Service's classification system. Each Service has developed its own ASVAB composites and has established minimum cut scores for each of its jobs or occupational areas on one or more of its composites to ensure a minimum level of aptitude for each job. Additionally, each Service uses aptitude scores to match people to jobs. However, the way in which this "match" is made and the type of information that goes into the "matching" process vary considerably by Service. The actual assignment of recruits to occupational areas or jobs is accomplished via computerized Person Job Match (PJM) algorithms. Each Service has its own algorithm, which reflects its current policies toward the relative priorities of filling jobs at any point in time.

With the exception of the Army, all Services use a two-stage PJM process. In the first stage (pre-enlistment), applicants are assigned to jobs, or to occupational areas, through a sequential processing system. [The Army assigns all applicants to jobs at the MEPS.] Most of the post-enlistment assignment systems operate in batch mode. Sequential processing refers to assigning one individual at a time to one of a number of jobs while simulating a batch processing environment. In batch processing, groups of recruits are assigned to jobs concurrently; characteristics of recruits can be compared directly in matching individuals to jobs. When recruits must be processed individually, no direct comparison group is available. Therefore, the pre-enlistment algorithms make comparisons based on a shadow population resembling the real one to which the present recruit belongs.

Summary. There are four major differences in selection and classification procedures across the Services. First, the Services use somewhat different cut scores on AFQT for selection. Second, the Services use somewhat different variables for classification (e.g., the Air Force is the only Service that uses a strength measure). Third, the Army assigns all applicants to jobs at the MEPS while the other Services use a two-tiered classification system, and fourth, the PJM systems used by the Services to classify recruits into jobs differ.

The Military Classification Environment

Classification systems serve a set of organizational goals within complex interacting limitations. The number and type of jobs currently available (and forecasted) and the number and qualifications of recruits currently available (and forecasted) play a key role in determining whether an eligible recruit will be assigned to a particular job or occupational area. A host of other parameters (e.g., organizational policy constraints) affect the assignment decision.

The Goals of Classification

Classification systems are implemented to achieve a particular goal, or set of goals. Some exemplary goals are to (a) maximize mean individual performance across jobs, (b) minimize attrition across all jobs, (c) maximize training success across all jobs, (d) minimize the number of "problem" employees across all jobs, or (e) maximize the utility, or value, of performance across jobs.

The Services' classification systems have some similarities and some differences in terms of the goals they serve. (Figure 1 lists the goals served by the four systems). The systems have two goals in common: (a) to maximize the utilization of training school vacancies across jobs, and (b) to ensure individuals meet minimum requirements of jobs. The Army's system has a goal that the Army calls the "quality goal"--to match the distribution of aptitude within jobs to a desired distribution. Specifically, the Army has established goals for AFQT categories within most MOS. These goals are used to ensure that a fixed percentage of high aptitude recruits are assigned to low complexity jobs, thus ensuring a sufficient source of trainers and noncommissioned officers in each specialty. The Air Force and Navy systems serve several goals and are similar to each other.

Assignment Goals	Air Force PROMIS	Army REQUEST	Marine Corps ARMS	Navy CLASP
Maximize utilization of training school vacancies across jobs	X	X	X	X
Ensure individuals meet minimum aptitude/physical requirements of jobs	X	X	X	X
Maximize the fit of individual aptitudes to job difficulty/ complexity	X			X
Maximize predicted training success	X			X
Maximize the fit of individual preferences to jobs/occupations	X		•	X
Match the distribution of aptitude within jobs to a desired distribution (to ensure distribution of quality across jobs)		X		

Although the ARMS System does not incorporate job preference information, job preference is used in Marine Corps assignments to start the ARMS system.

Figure 1. Goals Served by Each Service's Pre-Enlistment Assignment System

When we interviewed selection and classification experts for this project, we found that representatives from the Services agree that classification systems should "maximize aggregate predicted job performance across jobs"--a goal that is noticeably absent from the goals served by the current assignment systems. Two other goals used by the Air Force and the Navy might be thought of as proxies for maximizing predicted performance. Fitting individual aptitude to job difficulty or complexity, for example, is an indirect way of predicting success on the job from aptitude measures. The assumption underlying the aptitude/difficulty function is that high aptitude individuals will perform better in complex jobs than will low aptitude individuals. Similarly, fitting individual preferences to job/occupations is based on the assumption that matching individuals to their preferred assignments will increase job satisfaction and, perhaps, reduce attrition or

enhance job performance.

<u>The Future Demand for Recruits</u>. Changing missions and limited resources are likely to result in a different kind of future occupational structure. For example, DoD involvement in the war on drugs or in the defense of our borders against illegal alien entry may require more small plane pilots and small intervention units that operate autonomously. Also, in response to funding limitations, the Services are redesigning jobs to make the workflow more efficient. This means that the Services are headed toward more general jobs and fewer specializations. Such changes involve restructuring the workforce, or changing the demand side of the supply-demand equation.

<u>The Future Availablity of Recruits</u>. The applicant pool is the supply side of the assignment equation. As we approach the year 2000 workforce demographics are changing. The workforce is aging and will contain proportionally fewer young adults. There will be proportionally more women and minorities, particularly Hispanics. Ree and Earles (1991) applied demographic trend information to the 1980 ASVAB norming sample to estimate the effects of demographic change on the Air Force's applicant population. Their findings suggest that the numbers of young people in AFQT Categories I-IIIa will decrease. There is also concern that the workforce will (a) lack the skills and education needed to meet the demands of advanced technological jobs of tomorrow, and (b) lack English language proficiency and/or literacy.

<u>Constraints on Classification Systems</u>

Constraints are factors that limit the feasibility or usefulness of optimal classification strategies. In all organizations, the classification decision-making process must operate under one or more constraints (e.g., budget limitations, training seat availability, goals for specific subgroups, management priorities, applicant preferences). In general, the existence of constraints reduces potential gains from classification.

<u>Training Seat Availability</u>. Training seat availability is the single most important and influential constraint operating on military classification systems. Training numbers (how many people can be trained), timing (when training occurs), and priorities (the criticality of the organization's need to fill the job) enter all of the Services' allocation equations.

<u>Occupational Preference</u>. Occupational preference is a constraint on classification because it may limit optimal assignment based on aptitude measures and because applicants may not be equipped to make good occupational decisions; in turn, preferences may add error to the assignment. Occupational preference has assumed a larger role in military classification since the move to the All Volunteer Force. Individuals' preferences have to be considered, to some degree, in job assignments. Both the Air Force and the Navy include preference scores in their assignment systems. The Army does not incorporate preferences in the assignment system, although the Army, like the other Services, provides a list of jobs from which the applicant may choose (or, even choose not to enlist).

Minimum Aptitude/Physical Requirements. All Services impose minimum aptitude and physical (e.g., color vision, height) requirements for jobs, and the Air Force has the additional constraint of a minimal strength requirement for some of its jobs. The degree to which these minima constrain classification depends on the stringency of the cut scores and the distribution of the attribute in the recruit population. Higher cut scores place greater constraints on the system.

Minority/Nonminority Fill Rates. The Marine Corps and Navy impose minority/nonminority fill rates for some jobs. It is difficult to discern, however, how stringent these policies are or how much they constrain classification. Also, all Services adhere to combat exclusion laws (or policy in the case of the Army) prohibiting women from combat jobs. Combat exclusion laws and policies are currently under review and may change.

Organizational and Societal Constraints. Several other factors limit classification systems, although they do not appear in most algorithms. With the draw-down, resources have become even more of an issue. Diminished funding for personnel (testers, classifiers), places, and equipment puts real limitations on what is operationally feasible. Societal considerations and politics also affect military classification systems. The relationship between doing something for the good of society and the good of the force varies with political administrations and the environment. Administrations can communicate socially responsive messages through military policy (e.g., stay in school) or even suggest using the military to achieve social goals (e.g., upward mobility). Clearly, military selection and classification occur within the current socio-political environment; some limitations are imposed by society.

Research Objectives

Anticipated changes in the supply or demand for recruits or the constraints on classification systems affect future research needs. For example, the war on drugs and other mission changes may result in more special operations and low intensity warfare of short duration. The Services may need the capability to form small, quick-reaction teams of highly specialized personnel for small conflicts around the world. Such emphases imply that the Services may need to expand research on (a) the characteristics individuals need to perform these tasks effectively, and (b) team performance and how individuals contribute to the performance of the team. In all, the anticipated changes and other topics described by interviewees resulted in a list of 25 research objectives.

Camara, W. J., & Laurence, J. H. (1987). Military classification of high aptitude recruits. (FR-PRD-87-21). Alexandria, VA: Human Resources Research Organization.

Ree, M.J., & Earles, J.A. (1991). Estimates of available aptitude as a consequence of demographic change (AL-TP-1991-0019). Brooks, AFB, TX: Armstrong Laboratory.

Russell, T.L., Knapp, D.K., & Campbell, J.P. (1992). Building a joint-service clasificaiton research roadmap: Defining research objectives (HumRRO IR-PRD-92-10). Alexandria, VA: Human Resources Research Organization.

BUILDING A JOINT-SERVICE CLASSIFICATION RESEARCH ROADMAP: PREDICTOR-RELATED RESEARCH NEEDS

Douglas H. Reynolds, Teresa L. Russell
Human Resources Research Organization (HumRRO)
John P. Campbell
University of Minnesota and HumRRO

Military selection and classification researchers have had a continuing interest in the development and refinement of psychological measures that predict job performance. The currently operational selection and classification battery, the Armed Services Vocational Aptitude Battery (ASVAB), is a valid and fair measure of general cognitive ability (Welsh, Kucinkas, & Curran, 1990). However, the Services have an interest in building upon the test to improve its usefulness. Thus, the Services have been actively involved in research on supplements for the ASVAB as well as alternative predictors of performance. For example, the Enhanced Computer Administered Test battery (ECAT) has been developed to broaden the measurement of specific cognitive constructs that are not covered by the ASVAB. As a part of a Joint-Service effort to examine existing predictor research and formulate objectives for future research, we reviewed recent research on various predictor measures conducted by each of the Services.

We organized our review around two central questions: (a) what types of predictor measures are most likely to be useful (i.e., valid and fair) predictors of performance on military jobs, and (b) what research is needed before these measures can be considered for operational use. When conducting our review, we did not set out to find answers but rather to develop a set of objectives for future research. We proceeded by reviewing current research on the ASVAB and existing alternative measures, such as the ECAT. We also examined research on experimental predictors of cognitive abilities, psychomotor abilities, physical abilities, personality characteristics, interests, and biographical attributes. This paper discusses the major findings from the literature review. A full summary of our review is available in Russell, Reynolds, and Campbell, 1992.

Currently Available Predictors

The ASVAB appears to be a highly useful general purpose predictor. Research indicates that ASVAB subtests, composites, and the ASVAB general factor are valid predictors of training and job performance (e.g., Welsh et al., 1990). The ASVAB predicts training success in a host of schools, for a variety of jobs, and in all the Services. Job performance validity information is limited but what is available indicates that the ASVAB predicts performance of the technical aspects of jobs (e.g., hands-on tasks). Current efforts to improve the ASVAB is focusing on two major areas: (a) broadening its coverage of cognitive constructs and (b) reducing its adverse impact.

Research has indicated that some important cognitive constructs (e.g., visualization) are not assessed by the ASVAB (e.g., McBride, 1991). Additionally, some studies have noted sex and race differences on the measure (e.g., Peterson, Russell et al.,

1990). The Services recognized these deficiencies in the ASVAB when preparing the ECAT. During the course of our review, we examined the research to date on the ECAT measures. So far, several ECAT measures look like good candidates for inclusion in the ASVAB. With regard to spacial visualization, for example, the available data suggest that ECAT Assembling Objects test maybe worthy of consideration as a supplement to the ASVAB. The test has yielded small sex differences (relative to other spatial measures) in three large samples and has been a useful predictor in studies conducted by the Marine Corps as well as the Army (e.g., Peterson, Hough et al., 1990).

With the Joint-Service ECAT project, the Services are well on the way to identifying changes in new versions of the ASVAB. Short-term research projects do, however, continue to be necessary to identify the impact of removing specific subtests from the ASVAB and inserting new ones. These efforts are also underway in each of the Services.

New Predictors

In order to propose research objectives regarding "new predictors" (i.e., measures that are not refined enough for operational use), we examined research on cognitive, psychomotor, physical ability, personality, interest, and biographical measures. Currently, information about tests used by the Services is not easy to collect, and available information can be spotty. For example, race and sex differences are often not reported. The information that is available is inconsistent in format and is difficult to cumulate. Recomputations are often required so that results can be reported in a common framework. Also, there is no central resource where test information is available. Researchers planning to develop new tests or a battery of tests must do considerable "leg-work" (phone-calls, literature searches) to find out whether another Service is undertaking a similar effort or has such tests on hand. A Joint-Service Test Bank would maintain a data base of descriptive and psychometric test information for military research purposes. Thus, one proposal from our review is that a Joint-Service Test Bank be developed to enhance the accessibility of both officer and enlisted test information, to encourage experimentation across Service boundaries, and to build knowledge about current tests. In addition to this general proposal, we have generated specific proposals regarding new predictors that are related to each of the areas we reviewed.

Cognitive Predictors. Several new predictors from ongoing projects at the Air Force, Navy, and Army (e.g., Peterson, Hough et al., 1990), hold promise for supplementing future ASVABs. Research using available cognitive measures should be encouraged wherever possible. Doing so would not only reduce costs associated with test development but also enable a richer base of knowledge about tests to be build. Also, basic research on cognitive abilities is needed to identify abilities, enhance measurement, learn more about how abilities change over time or with practice, and to link information processing and traditional abilities domains. To this end, we propose two research objectives related to cognitive predictors: (a) include selected, already developed cognitive predictors in validation studies, across Services--to identify candidates for inclusion in future ASVABs; and (b) continue to research basic cognitive abilities measurement.

Psychomotor Predictors. The ECAT contains two tracking tests. The addition of these tests to the ASVAB would represent measurement of a new domain, and there is reason to expect that these psychomotor tests would supplement the validity of the ASVAB. However, the addition of both tests is probably not necessary. ECAT Tracking 1 and 2 have virtually identical items and are highly correlated with each other (Peterson, Russell et al., 1990). Also, before implementing psychomotor tests, the Services will need to decide how to deal with the large practice effects associated with them. It may be possible to set up testing practice facilities in the Military Entrance Processing Stations (MEPS) or in recruiting stations, so that applicants could complete practice test items. Alternatively, a number of practice items could be included on the tests.

Sex differences on the ECAT tracking measures are large. As long as these tests are used for selection and classification for combat jobs and combat jobs remain off-limits for women, this is a moot point. If, however, combat exclusion policies and laws are removed in the future, a number of issues arise. First, perhaps it will be more important to use psychomotor measures to make classification decisions because a wider range of individuals may be considered for combat jobs. Second, because the sex differences are so large, it will be necessary to show that psychomotor tests, if used, are based on real job requirements identified through job analyses. Otherwise, it could be alleged that the Services adopted such tests as a surrogate for combat exclusion policies/laws, since psychomotor measures would exclude women from these jobs. Thus, we propose two objectives regarding psychomotor test research: (a) if psychomotor tests are to be used, a mechanism for dealing with practice effects should be developed and researched; and (b) job analytic research should be conducted to demonstrate the job relatedness of psychomotor abilities.

Physical Abilities Predictors. It is reasonable to expect that physical abilities measures would supplement the ASVAB for the prediction of performance in physically demanding jobs. Also, taxonomies of physical abilities are now available and can facilitate generalizability of validation results from civilian jobs to the domain of military jobs, making research less costly and more efficient (e.g., Hogan, 1991). Therefore, physical abilities predictors are good candidates for inclusion in future testing efforts.

The issues involved in implementing physical abilities and psychomotor tests are similar. Specialized job analysis information would be needed to determine the physical and psychomotor requirements of the jobs. Both types of tests will yield some, if not a great deal of, adverse impact. In the same vein, the issues of if, how, and where to appropriately set cut-off scores for the tests used would need to be addressed. Other practical considerations include the cost of acquiring special equipment to conduct physical abilities and psychomotor testing and hiring/training test administrators to validly and reliably measure individuals. Physical space arrangements would also need to be made for operating and storing test equipment.

Despite these concerns, assessing the capacity of military applicants to handle physical tasks would appear to be fundamental to selecting individuals to perform in certain fields. Hogan (in press) has indicated that, because physical abilities tests have

been found to be valid predictors of job performance and are statistically independent, they provide incremental validity to the prediction of the criterion space. The capability, then, exists to further calculate and thereby improve upon the performance of those entering and working in positions that require physical effort. Thus, we add the following objectives regarding physical abilities research: (a) if physical abilities tests are to be used, establish a job analytic mechanism for demonstrating the job relatedness of physical abilities; (b) examine and estimate the logistical requirements associated with physical abilities and psychomotor test administration; and (c) identify physical abilities measures that are likely to be good predictors with minimal adverse impact.

Personality Predictors. Personality predictors are promising candidates as supplements to the cognitive measures traditionally used by the Services for several reasons. First, recent advances in the area of personality structure have led to new agreement on basic factors around which traits may be organized. These factors have helped researchers to be specific about the nature of the criterion relationships that may be expected for personality variables. Second, meta-analyses have shown personality variables to have consistent useful relationships with a variety of criteria. Research from the Army's Project A (e.g., Campbell & Zook, 1990) indicates that personality measures are good candidates as supplemental measures to existing and experimental cognitive tests, especially for the prediction of "will-do" criteria such as effort, leadership, and personal discipline, as well as training attrition. Third, personality measures appear to show fewer differences between races than do cognitive measures, and the differences that have been shown tend to favor minority respondents. Fourth, the Services have already developed some personality measures that appear to work well.

The primary issue regarding actual implementation of personality measures is the potential for fakability and coachability. Faking is possible on these measures, but it is also possible to detect faking in many cases. Further research is necessary to determine how to best reduce socially desirable responding and purposeful faking and how to deal with suspect response profiles. The conduct of a comprehensive review of the faking and social desirability literature would be an important step in organizing our knowledge in this important area. The literature we reviewed suggests that the possibility that faking may occur does not completely deplete the utility of personality measures. It is also possible that there are ways to prevent faking that have not been explored (e.g., giving periodic, tactful feedback on a computer-administered form). Thus, we propose one major objective regarding personality testing research: investigate fakability/coachability of personality measures, particularly how to prevent fakability/coachability and how to determine the impact of faking when it does occur.

Interest Measures. The Air Force and the Navy currently use individual information about job preferences in their classification process (Russell et al., 1992). It is possible that interest inventories would more accurately identify interests than the current methods where recruits rate occupational categories, especially since new recruits tend to be job-naive. Validation findings indicate that interest measures predict later occupational membership and job satisfaction; however interests do not appear to add

much in the prediction of job performance over that accounted for by cognitive and personality predictors. These findings suggest that interest measures may by more useful for classifying people into jobs rather than as selection measures. There are two major obstacles to the implementation of interest measures for this purpose, however: (a) possible adverse impact and (b) the effect of coaching. Thus two objectives for future research on interests are: (a) analyze adverse impact issues regarding interest measures, and (b) identify ways to prevent faking/coaching on interest inventories.

Biodata Predictors. Biodata are effective and valid predictors of a number of important criteria. Research has indicated that biodata validities can be made generalizable and stable (Rothstein, Schmidt, Erwin, Owens, & Sparks, 1990), thus these measures are worthy of continued consideration as supplements to cognitive predictors of military performance. There is also evidence that biodata may have incremental validity over cognitive measures, especially when predicting non-performance criteria such as attrition (e.g., Trent, in press). Biodata do not yield large differences between the races and evidence of differential validity is slight. Although biodata measures are possible to fake, research indicates that faking may not be prevalent. Finally, one additional strength of biodata is that some measures (e.g., the Educational and Biographical Information Survey) account for variability in attrition that has traditionally been predicted by educational attainment criteria. Educational credentials have come under fire because they restrict entrance to the military for identifiable groups of individuals (e.g., GED recipients). Biodata instruments provide a compensatory measure such that no one particular characteristic will be likely to exclude an individual. Thus, biodata may face less implementation resistance than other predictors of military adjustment.

Biodata measures are probably one of the best candidates for improving enlisted selection and classification. If biodata measures are made operational, it is critical to track their performance over time and maintain the instruments accordingly. This leads to another objective: examine ways of limiting and detecting faking in biodata measures, and continue research to determine the utility of biodata predictors.

Multi-Domain Research. In an earlier phase of our study, Russell, Knapp, and Campbell (1992) interviewed military personnel specialists regarding the future needs of the Services. Three major themes regarding future changes emerged from our data. First, it was noted that in the future the Services will move from highly specialized jobs to jobs with more generalized responsibilities. Second, the mission of the armed forces is changing from large scale operations to smaller scale intervention, and with this change comes an increased emphasis on smaller teams that may be deployed quickly. Third, technological advancement will continue to change the nature of military work.

Such trends will likely lead to increased job complexity, greater social interdependence, and cognitive ability requirements that are beyond our current measurement capability. This suggests that these jobs may require higher cognitive ability, but also that selection and classification researchers will need to investigate the predictive utility of the interactions between cognitive and dispositional characteristics, basic differences in motivational predisposition and social intelligence, and the

measurement of basic cognitive processes. This suggests another research objective: conduct basic predictor research that spans several predictor domains and recognizes interactions between the domains.

Adverse Impact and New Predictors

Several of the research objectives we developed deal with the general issue of expanding the predictor space to fulfill the twin goals of incrementing the prediction of performance and reducing adverse impact on protected groups. It is important to note that the latter goal is not only a function of the tests that are used, but also how they are used. For example, the Navy requires that females attain a higher AFQT score than males for entry to enlisted jobs (Russell, Knapp, & Campbell, 1992). This suggests one final objective: identify policies that reinforce adverse impact and recommend alternatives.

References

Campbell, J. P. and L. M. Zook (Eds.) (1990). Building and retaining the career force: New procedures for accessing and assigning Army enlisted personnel, Annual Report, 1990 Fiscal Year (ARI FR-PRD-90-6). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Hogan, J. A. (1991). Physical abilities. In M. D. Dunnette and L. M. Hough (Eds.), Handbook of industrial and organizational psychology (Vol 2., pp. 751-831). Palo Alto, CA: Consulting Psychologists Press.

Hogan, J. A. (in press). Theoretical and applied developments in models of individual differences: Physical abilities. Proceedings of the ARI Conference on Selection and Classification. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Horn, J. L. (1989). Cognitive diversity: A framework of learning. In P. L. Ackerman, R. J. Sternberg, & R. Glaser (Eds.), Learning and individual differences (pp. 61-116). New York: Freeman.

McBride, J. R. (1991, August). Content and structural comparisons of ASVAB and DAT. Paper presented at the meeting of the American Psychological Association, San Francisco, CA.

Peterson, N. G., Hough, L. M., Dunnette, M. D., Rosse, R. L., Houston, J. S., & Toquam, J. L. (1990). Project A: Specification of the predictor domain and development of new selection/classification tests. Personnel Psychology, 43, 247-276.

Peterson, N. G., Russell, T. L., Hallam, G., Hough, L. M., Owens-Kurtz, C., Gialluca, K., & Kerwin, K. (1990). Analysis of the experimental predictor battery: LV Sample. In J. P. Campbell and L. M. Zook (Eds.), Building and retaining the career force: New procedures for accessing and assigning Army enlisted personnel, Annual Report, 1990 Fiscal Year (ARI Technical Report 952). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Rothstein, H. R., Schmidt, F. L., Erwin, F. W., Owens, W. A., & Sparks, C. P. (1990). Biographical data in employment selection: Can validities be made generalizable? Journal of Applied Psychology, 75, 175-184.

Russell, T. L., Knapp, D. J., & Campbell, J. P. (1992). Building a Joint-Service classification research roadmap: Defining research objectives (HumRRO IR-PRD-92-10). Alexandria VA: Human Resources Research Organization.

Russell, T. L., Reynolds, D. R., & Campbell, J. P. (Eds.) (1992). Building a Joint-Service classification research roadmap: Individual differences measurement. (HumRRO IR-PRD) Alexandria VA: Human Resources Research Organization.

Trent, T. (in press). The Armed Services Applicant Profile (ASAP). In T. Trent & J. H. Laurence (Eds.), Adaptability screening for the Armed Forces. Washington, DC: Department of Defense.

Welsh, J. R., Jr., Kucinkas, S. K., & Curran, L. T. (1990). Armed Services Vocational Aptitude Battery (ASVAB): Integrative review of validity studies (AFHRL-TR-90-22). Brooks Air Force Base, TX: U.S. Air Force Human Resources Laboratory.

BUILDING A JOINT-SERVICE CLASSIFICATION RESEARCH ROADMAP: CRITERION-RELATED RESEARCH NEEDS

Deirdre J. Knapp and John P. Campbell
Human Resources Research Organization (HumRRO)

ABSTRACT: This paper reviews research and issues related to criterion measurement for enlisted personnel. The focus of the review is on the recently completed Joint-Service Job Performance Measurement (JPM) Project as this work represents the primary foundation for future research efforts in enlisted personnel criterion measurement. Our review strategy was grounded in a conceptual model of job performance supplemented with a taxonomy of measurement methods. From this framework, we review and discuss the criterion measurement philosophies of the Services and the potential contributions of specific measurement methods to future research.

The Joint-Service Job Performance Measurement Project

The Services took a large step forward in addressing the sparsity of research focusing on the "criterion problem" when they embarked upon the Joint-Service Job Performance Measurement/Enlistment Standards (JPM) project (Harris, 1987). The JPM project was initiated as a result of a 1980 Congressional mandate which directed the Services to demonstrate empirically that performance on the Armed Services Vocational Aptitude Battery (ASVAB) is predictive of performance on the job. As part of the JPM project initiative, each of the Services embarked on individual programs of performance measurement research which were coordinated through the Joint-Service Job Performance Measurement (JPM) working group.

At the outset of the performance measurement project, the JPM working group identified hands-on work sample tests as the benchmark job performance measurement method against which less costly "surrogate" measurement methods, such as performance ratings or written job knowledge tests, would be compared. Furthermore, each Service was tasked to measure job performance for a sample of jobs using hands-on tests as well as a least one specific type of surrogate measure.

JPM criterion measures were developed for approximately 33 military jobs. Hands-on tests were developed for all of the jobs; written knowledge tests and rating scales were developed for more than one-half of the jobs. In addition, simulations (e.g., interactive video tests) were developed for a subset of jobs, and archival indices of performance (e.g., training grades) were identified for many of the jobs. Using these instruments, data were collected on over 15,400 enlisted personnel (26,400 if the Army's longitudinal validation sample is included).

The Services have produced a massive amount of criterion measurement information over the past decade. The JPM data sets are large, both in terms of sample sizes and types of variables, and cover a wide variety of jobs. In addition to the measures and data generated, the lessons to be learned from the development, administration, and analysis of these measures are significant.

A Conceptual Model of Performance

The conceptual model of performance adopted in our review is that proposed by Campbell, McCloy, Oppler, and Sager (in press). In this model, performance is defined as behaviors or actions that are relevant to the organization's goals. Furthermore, a job is a complex activity, and for any job, there are a number of major performance components that are distinguishable in terms of their determinants and covariation patterns with other variables.

Determinants of Performance. Individual differences on a specific performance component are viewed as a function of three major determinants: (1) declarative knowledge, (2) procedural knowledge and skill, and (3) motivation (McCloy, 1990). Of course, performance differences can also be produced by situational effects such as quality of equipment or differences in the degree of external support across individuals. For purposes of selection and classification research, however, situational determinants such as these should be kept constant as much as possible.

Latent Structure of Performance. The model is hierarchical with eight performance components at the most general level. Across jobs, the eight dimensions have different patterns of sub-dimensions and their content varies. Further, any particular job might not include all eight dimensions. These dimensions can be labeled as follows:

1. Job specific task proficiency
2. Non-job-specific task proficiency
3. Written/oral communication task prof.
4. Demonstrating effort
5. Maintaining personal discipline
6. Facilitating peer/team performance
7. Supervision
8. Management/administration

Coverage of Performance Dimensions in JPM Project. When viewed from the perspective of this model, important differences across the Services in their approach to performance measurement become clear. The Army attempted to measure all dimensions of performance that were identified through task-based and behavior-based job analysis. This roughly corresponded to dimensions 1, 2, 4, 5, and 6 above. In contrast, the other Services focused almost exclusively on job/occupation specific task proficiency (i.e., dimension 1). Ratings tapping other aspects of performance collected by the Air Force and Navy were generally excluded from validation analyses.

So just how much of the criterion space needs to be covered? Consider that there are many reasons why an organization might wish to measure job performance, and the preferred measurement strategy will be dependent upon the nature of those goals. With regard to validation needs, the Services basically have a two-stage system that they must support in their research (Russell, Knapp, & Campbell, 1992). The first stage, selection, is used to determine if an individual will be able to meet general performance requirements imposed by the organization (e.g., willingness to work hard and stick with the job). The second stage, classification (placement), is used to determine the jobs in which the individual is likely to perform most successfully.

The Army research was designed to validate not only the ASVAB, but other experimental cognitive and noncognitive selection and classification measures as well. Some of these measures were likely to be useful more for selection than for classification. To provide a reasonable test of these measures' predictive value, therefore, the Army focused on multiple components of performance. The other Services, however, were primarily interested in evaluating the predictive validity (both in terms of selection and classification) of ASVAB, and in some cases, other cognitive ability measures. Such measures seem most important for predicting technical task proficiency.

The way in which non-job-specific task proficiency was handled across the different Services appears to have been determined primarily by differences in force management. The Army has a clearly defined set of tasks which are required of all soldiers, regardless of Military Occupational Specialty (MOS) or occupational field. Thus, the Soldier's Manual of Common Tasks forms the basis for dimension 2 of the Campbell et al. performance model. Although the other Services may have job requirements that are force-wide, they are not clearly delineated in any documents of which we are aware. Thus, they tested no tasks that obviously fit into dimension 2 of the performance model. There were cases, however, in which tasks were identified as being common to an occupation or job set. The Marine Corps studied several MOS within a single occupational field whereas the other Services generally tested only on job from a given occupational field. Thus, a set of tasks common to all infantry MOS (or all helicopter maintenance MOS) were identified, as were sets of tasks specific to each MOS in the relevant occupation (e.g., rifleman, mortarman). The Air Force jobs were sufficiently varied that this Service identified both Air Force Specialty-wide tasks and job-type-specific tasks for testing.

Determinants of Performance. One can conceptualize the difference between maximal and typical performance measures as reflecting the degree to which the motivational determinant of performance is allowed to operate. Maximal performance measures essentially hold motivation constant whereas typical performance measures do not. Large differences in maximal task proficiency and typical task proficiency have been found (Sackett, Zedeck, & Fogli, 1988). The Services largely cpted for the measurement of maximal task proficiency in the JPM project. To some extent, this was based on the conclusion that we do not know how to measure typical performance well enough to justify doing so (Wigdor & Green, 1991). In its effort to obtain measures of will-do components of performance, the Army captured elements of typical performance using an array of archival indices of performance. The ratings collected by all of the Services also provided information regarding typical performance.

We agree that the assessment of typical performance is difficult, and is only possible with a couple of measurement methods. The inclusion of typical performance information in a set of criterion measures, however, will permit a more accurate assessment of actual on-the-job performance. Furthermore, we expect that using exclusively maximal performance criteria will lead to overestimates of the operational predictive validity of ability-based predictors because they ignore the very real influence of motivation.

A Taxonomy of Measurement Methods

Our taxonomy of measurement methods distinguishes four major methods: (1) Performance (e.g., hands-on) tests which present task stimuli in a fairly realistic fashion using actual equipment or props; (2) Verbal (e.g., written, oral) tests which describe task stimuli using words rather than equipment or props; (3) Ratings which are evaluations of actual job performance; and (4) Archival records which are indices of actual performance that are available without the use of non-routine data collection activities. Figure 1 depicts this measurement method taxonomy, and lists specific measurement strategies within each category.

PERFORMANCE	VERBAL
Work Samples	Traditional multiple-choice Performance-based multiple-choice
Simulations • Computer/Visual/Audio Aids • Assessment Center Exercises	Written Essay Oral Interview Accomplishment records

RATINGS	ARCHIVAL
Task-based Dimensional/Job-specific Global/Force-wide	Training grades Supervisor ratings Promotion rate Rewards
Supervisor Peer Self	Disciplinary actions Production Indices Turnover

Figure 1. Summary of Measurement Methods

Discussion of Measurement Methods

Performance Tests. It is widely accepted that work samples and simulations are capable of providing valid and fair measures of performance. In terms of our performance model, a well-designed performance test will allow one to assess procedural knowledge and skill (McCloy, 1990). The disadvantages are widely known as well. Although it is not particularly expensive to develop work sample tests, administration costs, in terms of equipment, people, and time, are very high. In addition, time requirements mean that either a large amount of time needs to be devoted to testing (i.e., 1-2 days) or very few tasks can be covered. To the extent that relatively few tasks are tested, the content validity and reliability of the resulting measurement is a significant concern.

In addition to traditional hands-on tests, the Services experimented with oral "walk-through" performance testing (Air Force) and interactive video testing (Navy) within the context of the JPM project. Oral performance testing appeared to work fairly well. This measurement method is most useful, however, as a supplement rather than replacement of hands-on testing because oral performance tests are no less resource-intensive to develop or administer. Interactive video testing, however, has the potential for offering a more practical performance testing strategy. Such a measurement system could exhibit relatively high fidelity and good psychometric

characteristics in a self-contained transportable computerized package. Unfortunately, little has been published regarding the development and administration of the Navy simulations.

Verbal Tests. Verbal tests have been praised for their economy and convenience and disparaged for their lack of realism because they can only assess declarative knowledge. The increasing sophistication of written and oral test strategies, however, is allowing some headway on the realism issue. The written knowledge tests developed in JPM project were all "performance-based." Performance-based items make liberal use of figures and pictures to depict task stimuli, and focus on how a task is performed rather than why it is performed in a certain way. Items can also be written which pose complex technical or supervisory judgment problems (e.g., the Army's Situational Judgment Test designed to test first line supervisory skills). These tests are harder to develop because the development of an answer key requires expert judgment rather than reference to training documents or textbooks. Finally, we note that verbal tests can be used to cover a wide variety of tasks in less testing time and allow the depiction of many different task conditions. Unfortunately, the Services have not fully examined the utility of these tests given the JPM data, although some work has been reported. Given the general economy and feasibility of this measurement method, such examination should be conducted to the fullest extent that the data will allow.

Performance Ratings. Conceptually, ratings would be the ideal measurement method because they are intended to capture typical on-the-job performance which is determined by declarative knowledge, procedural knowledge and skill, and motivation. Unfortunately, it turns out that people are often not very good at making ratings, with the major problem being various types of criterion contamination. Although each of the Services collected ratings in the JPM project, the extensiveness of their efforts varied widely. The Army had a variety of rating instruments and collected ratings from two supervisors and three to four peers. In contrast, for its Infantry MOS, the Marine Corps collected ratings from a single supervisor on a two-item scale. As with the written tests, reported analyses of the JPM ratings have been very limited. Army analyses suggest, however, that carefully collected for-research-only ratings from multiple raters and rater types yield reliable and valid performance information (Pulakos & Borman, 1986). The feasibility of this measurement method, along with encouraging research of this nature, argue for continued consideration of its utility, especially when used in conjunction with other methods (e.g., verbal tests).

Archival Records. Indices such as turnover, disciplinary actions, awards, and promotion rate show considerable promise as supplemental criterion measures, but they do not provide job specific performance information. An exception is training grades. Training grades are problematic, however, because the quality of tests and the nature of the score distributions vary widely across jobs. Furthermore, most of the Services have had significant problems with the accuracy of the training data bases, and the Marine Corps no longer even maintains such a data base. We conclude that archival records are potentially quite useful for selection research, especially since some indices can capture important elements of typical performance. This source of performance information, however, is not likely to be often useful for classification.

Summary

Our review of performance measurement methods in view of the Campbell et al. performance model suggests that the choice of performance dimensions to measure must be determined by the specific goals of the research. Obviously, however, the more comprehensive the coverage, the greater the potential uses of the data beyond those which originally motivated the research. Furthermore, if reasonably valid indicators of typical performance can be incorporated into the criterion measure set, the resulting data will yield more complete validity estimates than those which will be obtained with maximal measures used alone. Speaking more generally, whether one is trying to measure one performance dimension or several, every effort should be made to use multiple measurement methods. This multiple measurement method strategy should enhance measurement reliability and account for more performance determinants.

Space limitations prevent a comprehensive reporting of the Roadmap project's review of criterion measurement issues here. (See Knapp & Campbell, 1992 for the complete review.) Instead, we have tried to provide a description of the approach we took to this review and a sampling of some of our observations. An overriding conclusion, which may not be clear from the information presented in this paper, is that existing JPM data can be used to examine many outstanding criterion measurement issues. The utility of different combinations of non-hands-on tests for representing job performance (e.g., can a combination of written test scores and supervisor ratings yield useful criterion measures) and the adequacy of different task sampling strategies (e.g., does the job element portion of the Marine Corps sampling strategy improve the mix of tasks selected for testing) are just two examples of issues that can be examined further with existing data. Before additional resources are directed toward the development and administration of new criterion measures, therefore, we believe that it is in the Service's best interests to examine more fully the instruments and data already in their possession.

References

Campbell, J.P., McCloy, R.A., Oppler, S.H., & Sager, C.E. (in press). In N. Schmitt & W.C. Borman (Eds.) Frontiers in industrial/organizational psychology: Personnel selection. San Francisco: Jossey-Bass.

Harris, D.A. (1987, March). Job performance measurement and the joint-service project: An overview. In Proceedings of the Department of Defense/Educational Testing Service Conference on Job Performance Measurement Technologies. San Diego, CA.

Knapp, D.J., & Campbell, J.P. (1992). Building a joint-service classification research roadmap: Criterion-related issues (HumRRO IR-PRD-92-xx). Work performed under Contract No. F33615-91-C-0015 with the Air Force Armstrong Laboratory. Alexandria, VA: Human Resources Research Organization.

McCloy, R.A. (1990). A new model of job performance: An integration of measurement, prediction, and theory. Unpublished Ph.D. dissertation. University of Minnesota, Minneapolis, MN.

Pulakos, E.D., & Borman, W.C. (Eds.) (1986). Development and field test of Army-wide rating scales and the rater orientation and training program (ARI Technical Report 716).

Russell, T.L., Knapp, D.J., & Campbell, J.P. (1992). Building a joint-serviceclassification research roadmap: Defining research objectives (HumRRO IR-PRD-92-10). Work performed under Contract No. F33615-91-C-0015 with the Air Force Armstrong Laboratory. Alexandria, VA: Human Resources Research Organization.

Sackett, P.R., Zedeck, S., & Fogli, L. (1988). Relations between measures of typical and maximum job performance. Journal of Applied Psychology, 73, 482-486.

Wigdor, A.K., & Green, B.F. (Eds.) (1991). Performance assessment for the workplace (Vol. I). Washington, DC: National Academy Press.

Building a Joint-Service Classification Research Roadmap: Classification Research Objectives

John P. Campbell
University of Minnesota and Human Resources Research Organization

Teresa L. Russell, Deirdre J. Knapp, and Douglas H. Reynolds
Human Resources Research Organization

This paper describes an effort to conduct a needs analysis for personnel classification research and to identify the future research objectives that the key scientific and management personnel in the various Services and the Department of Defense view as the most critical. Subsequent parts of the project will attempt to identify the most relevant existing literature that pertains to these objectives and then generate an ordered sequence of future research activities that would address the most relevant issues in a profitable way.

There were essentially two major steps. The first consisted of a series of interviews with key personnel to identify the full array of classification research objectives that are potentially relevant for maximizing classification effectiveness across the services during the coming decades, given the changes that are taking place in mission, technology, and structure. The second step was to conduct a more formal survey and ask the personnel to prioritize the full list of objectives.

Method: Part I

After generating and pretesting an interview protocol we interviewed 43 individuals from the Armstrong Laboratory, the Army Research Institute (ARI), the Navy Personnel Research and Development Center (NPRDC), the Center for Naval Analyses (CNA), the Military Accession Policy Working Group (MAPWG), the Defense Manpower Data Center (DMDC), and the Office of the Assistant Secretary of Defense for Force Management and Personnel (OASD-FM&P). The sample consisted of the professional, scientific, and management personnel from these organizations who are most concerned with selection and classification issues.

The interviews were designed to (1) brief the participants on the Roadmap project; (2) ask for opinions about the most appropriate objectives for classification research over the next ten to 15 years, and (3) obtain information about any research currently being directed toward such objectives.

After the first set of interviews, we prepared an initial list of potential objectives that became part of the protocol for the next set of interviews. With such subsequent interview the list of objectives was revised on the basis of the new input. In short, the final description of potential objectives evolved iteratively over the course of all the interviews. The objectives were then clustered into categories based on their content similarity. The categories are shown in Figure 1.

A. Improve classification efficiency by restructuring the decision sequence and redefining the decision outcomes.

1. Investigate job clustering methods to improve potential for classification among appropriate job clusters rather than among individual jobs.

2. Develop/evaluate alternative paradigms for the selection/classification decision sequence (e.g., manipulate timing of classification decisions; make multi-level or multi-tiered classification decisions).

B. Improve classification efficiency by improving predictor and criterion measurement.

3. Investigate job analysis methods that more adequately capture nonobservable job requirements for high level performance (e.g., cognitive task analysis).

4. Design and evaluate job analysis methods that yield task to aptitude linkages, within defined task and aptitude taxonomies, so that aptitude requirements for jobs are readily and systematically defined.

5. Design and evaluate job analysis methods that identify the major contributions of individual performance to unit performance.

6. Investigate ways to improve the classification utility of existing predictors (e.g., revisions of weighted composites for ASVAB).

7. Determine which existing (but not implemented) predictors are most useful for classification purposes.

8. Develop and evaluate measures of new predictors likely to be useful for classification purposes.

9. Investigate optimal strategies for incorporating predictor information into the assignment decision (e.g., alternatives for developing and using composites).

10. Investigate criterion issues (e.g., How does the type of criterion used in validation affect estimates of classification efficiency and, ultimately, classification decisions? What is the appropriate criterion?).

11. Investigate optimal selection and classification strategies that maximize the contribution of individual performance to unit performance.

C. Improve classification efficiency by improving the operational assignment system.

12. Build an optimal assignment model that minimizes the impact of constraints on optimal assignment (e.g., "look-ahead" vs. strictly sequential processing to reduce impact of training slot availability).

13. Increase flexibility of assignment system (e.g., its responsiveness to supply and demand fluctuations).

Figure 1. Classification Research Objectives

14. Investigate ways to maximize the influence of predicted performance in the assignment system (e.g., improve composite standard setting procedures; incorporate predicted performance into assignment algorithm).

D. Evaluate alternative strategies for improving classification equity/fairness (i.e., minimize adverse impact; minimize predictive bias; maximize fairness of the global impact of classification decisions throughout the selection and classification system).

15. Evaluate alternative fairness models in terms of their effects on selection/classification outcomes across subgroups.

16. Develop and evaluate extended models of fairness/equity issues by mapping out consequences of classification decisions at various stages in the selection and classification process.

17. Identify and/or develop classification measures that minimize adverse impact and/or predictive bias.

18. Investigate alternative selection and classification criterion measures in terms of their relative construct validity and susceptibility to subgroup bias.

E. Evaluate and develop alternative classification models (i.e., generalizability, cost-effectiveness).

19. Improve classification efficiency by improving strategies to generalize classification research findings across jobs and military populations.

20. Develop and evaluate alternative strategies and models for estimating the cost-effectiveness of an alternative classification system in terms of reduced training costs, reduced attrition, dollars, etc.

F. The following items are not research objectives per se; however, they are objectives of those who are responsible for interfacing with the user community:

21. Identify classification system decisions traditionally driven by policy directives rather than psychological research findings (e.g., exclusion of women from combat jobs, physical or moral standards).

22. Develop user-friendly operational assignment systems.

23. Establish mechanisms for collecting necessary research data that minimize impact on operational systems.

24. Establish mechanisms for better communication with classification system users.

25. Develop ways to use the classification system to facilitate lateral career moves.

Figure 1. Classification Research Objectives (Continued)

Method: Part II

The list of potential objectives shown in Figure 1 became the content of a survey questionnaire that attempted to elicit expert judgments about the criticality of each objective as a future research priority.

Via a mailed survey, 32 individuals representing the Armstrong Laboratory, ARI, CNA, NPRDC, DMDC, and OASD provided comments on the objectives and made two judgments about each one. First, each respondent was asked to consider the relative importance of each research objective for the effective accomplishment of his/her organization's mission, using the following scale:

0 = Not at all important/relevant
1 = Unimportant relative to other objectives
2 = Minor in importance relative to other objectives
3 = Important relative to other objectives
4 = More important than the other objectives
5 = One of the most important objectives

Second, each respondent was asked to estimate the urgency of his/her organization's need for addressing each objective, using the following scale:

1 = Long Range - Objective must be addressed within the next decade
2 = Urgent - Objective must be address within the next 5 years.
3 = Extremely Urgent - Objective must be addressed within the next 3 years.

Twenty-seven completed surveys were returned in time for analysis. The importance and urgency score were combined multiplicatively to form a criticality index (i.e., criticality = importance x urgency) with scores that ranged from 0 to 15.

Intraclass correlations were computed as an index of interrater agreement within each organization for the importance and criticality ratings. Mean ratings on each index for each objective were computed for the total sample and for each organization. The intraclass correlations was also used to index the level of agreement between organizations for the average (across raters within each organization) importance and criticality scores.

Results

Interrater agreement

The interrater agreement results are shown in Table 1. The levels of agreement between pairs of individual raters (i.e., k =1) are generally low, but there is also variability across organizations. The reliability of the mean ratings within organizations varies considerably across organizations and it is not simply correspondent with the number of respondents. Instead it seems to reflect the diversity among the individuals

within each organization. That is, the greater the diversity in the positions occupied by the respondents, the less the agreement in importance and criticality ratings. Keep in mind that disagreement in ratings is not necessarily bad. In fact, a future internal discussion of the reasons for such disagreements might be a very fruitful exercise.

Table 1

Estimates of Interrater Agreement of Importance Ratings and Criticality Scores[1]

		Importance Rating		Criticality Score	
	N	K=1	K=N	K=1	K=N
Air Force	9	.04	.27	.04	.25
Army	3	.16	.37	.30	.56
CNA	3	.30	.57	.45	.71
Navy	8	.12	.51	.17	.62
OASD/DMDC	4	.11	.32	.06	.19
Mean[2]	5	.27	.65	.28	.66

[1]Intraclass Correlation Coefficients (ICCs) were computed using Shrout and Fleiss (1979) formulas. K=number of raters.
[2]The profile of within-organization means. N=5 because there were five organizations.

Mean criticality ratings

Several objectives consistently received high criticality scores across organizations. Figure 2 shows mean criticality scores computed across organizations. Objectives 7. ("Determine which existing, but not implemented, predictors are most useful for classification purposes") and 10. ("Investigate criterion issues") were consistently rated as having high criticality. In most organizations, the predictor and criterion-related research objectives were also judged to have high criticality (Objectives 6 through 10, 15, and 17).

The objectives that were consistently rated low in criticality had to do primarily with job analysis issues (3,4) and some of the operational personnel management concerns (21, 22, 25).

Although the data for each organization are not shown, the objectives showing the most variation in criticality ratings across organizations had to do with the relative emphasis to be given to (a) better use of existing predictor measures vs. (b) utilization of new predictors that are already developed but not yet implemented vs. (c) development of additional new predictors. There is also some disagreement about whether new (as opposed to existing) models of fairness should be investigated.

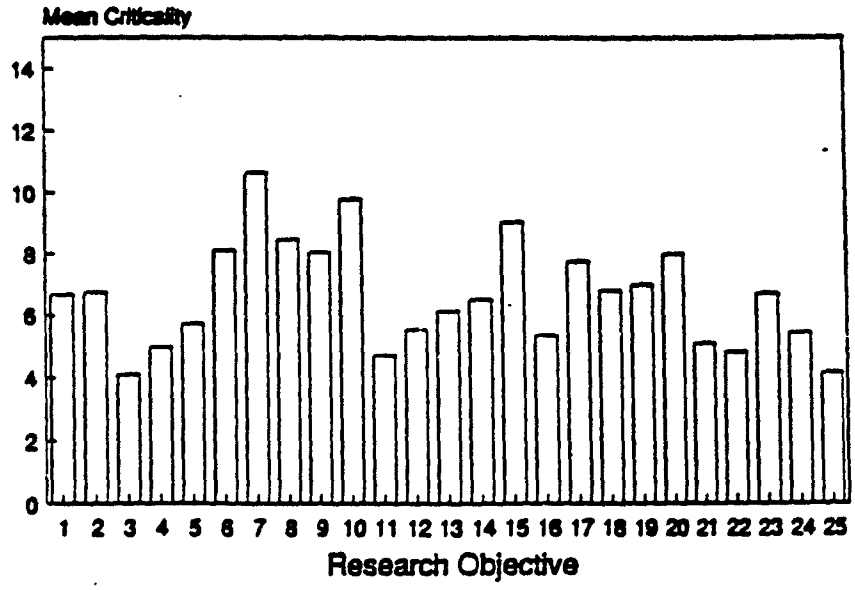


Figure 2. Overall Criticality Scores (means of within-organization means)

In general, the variation across objectives within organizations tends to reflect the long standing interests of the organization, which is to be expected. We indeed do tend to work on things that we judge to be important.

Implications

One principal implication of these results is that almost everyone thinks the greatest research payoff will continue to be in predictor development, criterion development, and a more meaningful structuring of jobs into homogeneous families that will facilitate classification. New job analytic techniques or more sophisticated quantitative models are given a relatively low criticality. That is, it is the message and not the medium.

A second implication, related to the above is that the greatest disagreement among the Services is in exactly where to put the emphasis on predictor development. This leads to perhaps the most important implication of all. It would be desirable to use data like these as a starting point for a cooperative, intense, and ongoing discussion across all the Services about where to allocate research resources so as to maximize the collective gain from better personnel assignments.

THE ASVAB TEST CONTENT SPECIFICATIONS

John A. Harris
Defense Manpower Data Center

Introduction

In the course of developing achievement and aptitude tests over the past fifty years, accepted procedures for ensuring content validity have been established. They include specifying the domain of knowledge to be covered by the test, identifying measurable content, and developing content taxonomies to ensure appropriate content coverage. All these procedures assume that the test builder is beginning from scratch to construct a new test battery. What is more common in the real world of published achievement and aptitude tests is the attempt to revise and update an existing test battery, often without changing the basic statistical properties of the previous edition. Additional issues need to be addressed if the test has been standardized and new norms are not going to be obtained on the revised test.

The *Armed Services Vocational Aptitude Battery* (ASVAB) is a battery of ten subtests used for selection and placement in the military services. A student edition is used as a career counseling tool in high schools, as well as for qualifying for enlistment in the armed services. The enlistment version has six forms, and the student edition has four forms, for a total of ten forms in use at any one time. One form of the ASVAB was normed in 1980 and all subsequent forms have been equated to that form (the Reference Form). Comparability of forms is essential so that applicants are treated equally - regardless of when they are tested or which form is administered. In addition, the services need to know that similar numbers of applicants will be eligible for enlistment and that longitudinal studies can be reliably conducted for reporting to congress and the public.

Issues

Since the norming of the Reference Form in 1980, new forms have been developed by making them strictly parallel to the Reference Form. That is, items have been written to match on a one-to-one basis with Reference Form items. Items were matched on the basis of difficulty (P-values) and bi-serial correlations. Content was specified in terms of broad domains, but for all practical purposes, the content duplicated the Reference Form with only minor variations. Over a period of years, this has created several serious problems.

1. Since most of the items appearing in the Reference Form were written in the 1960's and 1970's, some of the content has become obsolete or is now inaccurate.

2. Since the domains of content were not carefully specified, items often covered a very narrow range of content, omitting some broad and important areas and often focussing on very specific bits of information.

3. In a number of tests, the items measure only definitions, recall of facts, or simple knowledge. The higher-order thinking skills of application, inference, and analysis are not measured.

4. Sensitivity to gender bias was not as emphasized when the Reference Form was assembled as it is today. Awareness of ethnic and gender bias, as well as the representation of different groups, needed more attention.

In addition to these general concerns there were specific criticisms of some tests which also needed to be addressed.

1. The Word Knowledge test was too easy with an average P-value (proportion passing the item) of about .81 for the Reference Form and all subsequent forms.

2. The Paragraph Comprehension test was relatively inefficient, with as few as one or two questions per passage for some passages. Many reviewers felt that to have slightly longer passages and to write more items per passage would make the test more valid.

3. The General Science test has relatively fewer items (approximately 12%-15%) measuring the earth and space sciences, compared to the life and physical sciences. The reason for this is that historically the earth sciences have shown differential proportion by gender. with females at a disadvantage relative to males. With the new awareness of and interest in environmental and ecological science, these differences may no longer be valid.

Procedures

When the Defense Manpower Data Center (DMDC) assumed responsibility for the development and publication of ASVAB, items for the new operational forms 20, 21, and 22 had already been developed. These forms, scheduled to be operational in the fall of 1993, were developed under the old model; that is, items were matched on a one-to-one basis with the Reference Form, normed in 1980. The first opportunity that DMDC had to develop items for new forms was in the summer of 1991 when item writing was done for the next set of student test forms (Forms 23 and 24), as well as additional item development for the newly conceptualized ASVAB item bank.

The Summer 1991 project involved writing items for that portion of the ASVAB known as the Armed Forces Qualification Test (AFQT). The AFQT contains four of the ASVAB subtests: Word Knowledge, Paragraph Comprehension, Arithmetic Reasoning, and Mathematics Knowledge. These four subtests provide a general measure of cognitive or academic ability (or "g") and are very similar to most academic tests of aptitude and achievement. Because of the author's extensive experience in developing academic achievement and aptitude tests, a content taxonomy based on this framework was used as a starting point. This taxonomy covered the domains and categories normally measured in the most commonly used achievement and aptitude tests.

Items from the Reference Test were then mapped onto this objective structure and the results analyzed. Critical judgments were made in those instances where there were objectives not measured by the Reference Form as well as items on the Reference Form for which there were no objectives. Additional item content mappings were done on each set of objectives, using items from the new student forms (Forms 18 and 19) and the new operational forms (Forms 20, 21, and 22) which had been previously developed and were in the equating and tryout stages, respectively. These forms, similar to, but not identical to, the Reference Form permitted us to evaluate the degree to which new forms varied in content from the Reference Form. A number of factors were considered and iterations done to arrive at the final item-by-objective taxonomy.

One major consideration was the need to feel confident that the new forms could be

equated to the Reference Form until a renorming could be undertaken. New statistical methodologies in test development and equating have provided more flexibility and reliability in assuring test comparability. A description of these methodologies and how they are employed in the development of the ASVAB is provided in Bloxom and McCully (1992). These procedures, combined with editorial and content review, resulted in a new set of content taxonomies and item specifications for each test which was used by item writers as the basis for developing new sets of items. Table 1 shows the content taxonomy for Paragraph Comprehension. Table 1 also shows the difference in what is measured by, and the relative emphasis between, the Reference Form and the new forms.

The problem of passage efficiency was addressed by writing passages which were approximately one-third longer and planning to have 4 or 5 items per passage in the final selection. In order to increase the difficulty of the Word Knowledge test, all words tested in the tryout of Word Knowledge, as well as the words in the Reference Form, were graded according to difficulty. Each word was assigned a grade level according to two commonly used word lists.(Taylor et al, *EDL Core Vocabularies,* 1979; Dale and O'Rourke, *Living Word Vocabulary,* 1981)

The average grade level for the Word Knowledge test in the Reference Form turned out to be 8.2. To increase the difficulty of new editions the average grade level of the words was targeted at 9.2. It was estimated that this would lower the average P-Value from .81 to .75. The analysis of the item tryout data is still in progress, but it appears that the difficulty has been raised by about the desired amount.

We are currently involved in writing items for the four technical subtests of the ASVAB. They are General Science, Auto and Shop, Mechanical Comprehension, and Electronics Information. (Note: The other two ASVAB subtests are speeded clerical tests and are computer generated.) Unlike the AFQT subtests, the technical tests are designed to measure more specific abilities. One persistent criticism of the technical subtests is that they only measure basic knowledge, such as definitions and identification. It was felt that these tests, in order to be more valid measures of aptitude, should contain more items measuring higher level thinking skills.

Therefore, in addition to developing new content taxonomies similar to the AFQT tests, the specifications called for items to be written to three process categories: Knowledge, Application, and Analysis. Knowledge items test the ability to recognize, recall, define, locate, or identify. An item is an application item if the test-taker has to solve a problem, know how to use something or know how it works, or understand a concept or principle. To solve Analysis items, the test-taker must evaluate, infer, generalize, conclude, or interpret. Table 2 shows an example of a content-by-process taxonomy for General Science. It also includes a comparison of the item coverage on the Reference Form, as well as the new target distributions. Item development for the technical subtests is just getting under way; therefore, we will not have data to evaluate how these items worked until the Spring of 1993.

Summary

In the absence of renorming, the process of improving the content validity of the ASVAB is evolutionary. Since each subsequent edition of the test must be equated to the Reference Form, the last normed edition, changes to the test must not be so different that the forms can not be equated. By carefully defining the domain of content to be tested, we believe that desired improvements to the test can be implemented while retaining the statistical integrity of the battery.

Reference

Bloxom, B. & McCully, R. (1992): *Initial Operational Test and Evaluation of Forms 18 and 19 of the Armed Services Vocational Aptitude Battery.* (DMDC Technical Report 92-001). Monterey, CA: Defense Manpower Data Center.

Table 1

CONTENT STRUCTURE - PARAGRAPH COMPREHENSION

Content	No. Items in Test	No. Items Ref. Form
A. Literal Comprehension	**7**	**(9)**
1. Identify stated facts	3	(5)
2. Identify reworded facts	3	(4)
3. Determine sequence of events	1	
B. Inferential/Critical Comprehension	**8**	**(6)**
1. Draw conclusions	2	(3)
2. Identify main idea	2	(1)
3. Determine author's purpose	2	(1)
4. Determine author's tone/ mood	1	(1)
5. Identify style and technique	1	
TOTAL	**15**	**15**

Table 2

CONTENT STRUCTURE - GENERAL SCIENCE

Content	*Process*			
	No. Items	KNOWLEDGE	APPLICATION	ANALYSIS
A. Life Science	**12**	**6**	**3**	**3**
1. Botany		(3)		
2. Zoology		(3)		
3. Anatomy & Physiology		(3)	(3)	
4. Evolution				
B. Physical Science	**10**	**5**	**3**	**2**
1. Force/motion		(1)	(1)	
2. Energy		(3)	(2)	
3. Fluids, gases		(1)		
4. Atomic structure				
5. Chemistry		(2)	(1)	
C. Earth/Space Science	**3**	**1**	**1**	**1**
1. Astronomy		(2)		
2. Geology				
3. Meteorology				
4. Oceanography				
TOTAL	**25**	**12**	**7**	**6**

NOTE: Bold numbers represent number of items per form.
Numbers in ( ) represent classification of Form 8a (Reference Form)

TEST ACCURACY SPECIFICATION

Lauress L. Wise
Defense Manpower Data Center

Introduction

There is a general belief that the scales for some ASVAB subtests may be non-optimal in that they are based on items with an inappropriate distribution of difficulties (e.g., too many easy WK items). The concern is that the resulting scores may not be sufficiently accurate at some points in the scale and may be more accurate than is required at other points.

As part of a general review of the ASVAB, DMDC is considering the item difficulty targets used in constructing new forms of the ASVAB. Two questions are under consideration. First, what should the target distribution of item difficulties be for each subtest? Second, how should new forms be constructed to assure adequate adherence to these targets?

In the past, the primary strategy for building essentially equivalent forms has been to match item difficulties to the reference form on an item-by-item basis. This procedure has generally been *sufficient* to produce new forms close enough in overall difficulty to the reference form so as to allow reasonable score equivalence through equating. The procedure places severe limitations on item development, however, as many good items cannot be used simply because they do not happen to match the difficulty of a reference form item. Item-by-item matching is not a *necessary* procedure. It is possible to construct forms of equivalent difficulty by matching at the level of the overall difficulty distributions.

Underlying the general questions about item difficulty is the general issue of how accurate the scores generated for each subtest in the form should be. This is a policy question, involving tradeoffs between the variable costs of test development and administration related to precision and the benefits from more accurate estimates of the underlying abilities. The purpose of this paper is to describe our strategy for linking policy judgments about test accuracy to the procedures used in assembling forms. The two general issues addressed are how best to portray the accuracy of test forms and, given this portrayal, how to estimate the accuracy of test forms.

Background

Classical test theory (CTT) (e.g., Lord & Novick, 1968) describes test accuracy in terms of the reliability coefficient, an estimate of the correlation of scores from two parallel forms. This approach assumes that error of measurement is constant throughout the measurement scale. This assumption may not seem tenable, as it would seem that a test form with mostly easy items would be more accurate at the low end of the ability scale than at the high end. One must remember, however, that classical test theory was designed for use with a number right score. At every point in the scale, one unit corresponds to one more item correct so that the homogeneity of error assumptions may not be as unreasonable as they appear. Nonetheless, several efforts have been made to estimate errors of measurement for specific number correct score levels (Qualls-Payne, 1992; Feldt, Steffen, & Gupta, 1985).

In general, a number right score may not be the best metric for consideration of difficulty targets. The relationship between an examinee's true ability and his or her number

right score depends very heavily on the difficulty of each of the items in the test form. Two forms with different item difficulty distributions will have different number correct score distributions for any given sample or population. In the ASVAB program, standardized subtest scores are used as the basis for forming composites, and for our most important composite, the Armed Forces Qualifying Test (AFQT) composite, we use percentile scores in making selection decisions.

Item response theory (IRT) has been advanced as an alternative to CTT in part to counter scale-constancy issues that arise with use of a number right scale. Latent ability is scaled so that the regression of each item score on the underlying ability follows a fixed functional form, usually a normal ogive or three-parameter logistic (3PL) function. Using IRT models, it is possible to estimate the accuracy of scores estimated for individuals at a given underlying ability level as a function of characteristics of the items used in the measurement. Thus, accuracy is viewed to vary across the measurement scale and can be estimated from item parameters (Lord 1980, Lord & Novick, 1968). Lord (1984) provided an approach for estimating score accuracy when scores are based on the number of correct responses rather than a direct (maximum likelihood or Baysian) estimate of underlying ability.

Score and Accuracy Metrics

In describing test form accuracy and setting accuracy standards, two critical questions are the metric used to describe score levels and the metric used portray accuracy. With respect to score metrics, unfortunately, neither the IRT theta metric nor the number correct metric is used in making personnel decisions. The most important metric is the Youth Population Percentile Metric used with AFQT. Similar percentile metrics are also used by the Air Force with their four aptitude composites. The other Services use sums of standardized (in the Youth Population) subtest scores. For the reference form, the standardized scales are essentially linear transformations of the number correct scores, but for more current forms a nonlinear equating transformation has been introduced in converting from number correct to standardized subtest scores.

The youth population percentile metric has been selected for portraying accuracy for two reasons. First, this is the metric used in the general determination of qualification. A second reason is that accuracy judgments should be linked to some population distribution. We should be relatively unconcerned about accuracy at points in the scale where there are few individuals to be evaluated and much more concerned at those points where many examinees will score. For the percentile metric, the relative number of examinees scoring at each point is essentially the same. (About two percent of the relevant population will score within one point of any given level.)

Given the choice of the percentile metric for describing examinee abilities, what metric should be used to describe accuracy? Typically, the standard error defined as the expected standard deviation across parallel forms (overall or at particular score levels) is used as a measure of accuracy. Alternatively, the distance between specified percentile points (confidence bound cutoffs) in the conditional distribution of observed scores given a "true" score might be used as the measure of accuracy.

We are currently pursuing a different metric for describing accuracy. The primary use of the test scores is to classify applicants, dichotomously, as either qualified or not qualified (overall or for a particular job). Consequently, we are using classification error rates as the measure of score accuracy. What proportion of examinees will be incorrectly classified, either as qualified when they are not (false positives) or as unqualified when they actually are (false

negatives). A classification error rate metric communicates the operational impact of score accuracy and may be more appropriate than standard error measures when communicating with policy makers.

One other issue concerning metrics is whether to consider subtest scores or composite scores. Composite scores are actually used in making personnel decisions. On the other hand, we assemble items for each subtest separately. Having to consider all of the consequences of item selection for all the different uses of each subtest would be complex. Further, composite definitions change over time and future changes are generally not known when new forms are assembled. Consequently, we propose to consider classification accuracy at the subtest level even though operational use is at the composite level. One other reason for controlling test difficulty at the subtest level is that the IRT model used in estimating accuracy assumes unidimensionality (local independence). This assumption is at least marginally tenable at the subtest level; it is not tenable at the composite level.

Proposed Approach

As described above, we are using classification error rates for defining/projecting score accuracy. We plan the following general procedure for reviewing current subtest accuracy profiles and for setting new targets as appropriate: (1) develop accuracy profiles for current forms; (2) use expert judgment to review/revised accuracy goals for new forms at the key points (ranges) on the target scale; (3) use IRT analyses to calibrate new items, adjust for differences between the tryout sample and the Youth Population, and estimate classification error rates for trial forms during form assembly; (4) develop preliminary tolerances for compliance with accuracy targets; and (5) check the initial form accuracy profiles against revised accuracy profiles computed on operational samples during formal equating and revise targets/tolerances as required.

The remainder of this paper presents details of the procedure proposed for obtaining item parameter estimates and using them to generate accuracy profiles for actual and potential forms. These procedures are illustrated with analyses of data from the Profile of American Youth Study.

Samples

The Profile of American Youth Study (OASD, 1982) provided the basis for the current ASVAB norms. It involved administration of the ASVAB reference form to a complex sample of approximately 12,000 youths. We drew a systematic sample of 4,000 cases from the data files using sampling probabilities that were inversely proportional to their current sampling weight. The overall selection probability was thus the original selection probability (the inverse of the sampling weight) times the probability of being selected for this new subsample (the weight times a constant). The composite probability was thus a constant, and the data could be analyzed without having to use case weights.

We next divided the 4,000 case sample into two 2,000 case samples (alternating in order of selection into the 4,000 case sample) for cross-validation purposes (and because we were using a PC version of BILOG to get item parameter estimates). The result of all of these machinations was two 2,000 case samples that were each representative of the entire youth population without having to use differential case weights.

For illustrative purposes, we examined the WK and GS subtests. WK is notorious for

having an abundance of relatively easy items, while GS is more balanced with respect to item difficulty.

Methods

IRT parameter estimates. We obtained item parameter estimates for each of the two subtests in each of the two 2,000 case samples. The BILOG program was used with options specifying floating priors for the slope and asymptote (a and c) parameters and no prior for the threshold (b) parameters. If this were not a strictly representative sample from the Reference Population (RP), we would have to adjust the item parameter estimates for differences between the calibration sample and the RP. Typically, the reference form is administered to a sample that is randomly equivalent to the sample used to calibrate new items. Differences in reference form item parameter estimates from the youth population sample and the new sample provide the basis for translating the new item parameter estimates back onto the reference population theta scale. One approach, for example, is to find the linear transformation that minimizes the (weighted) average squared difference in the test characteristics curves based on the original calibration and the rescaled new estimates (Lord & Stocking, 1983). Alternatively, the differences to be minimized may be expressed relative to the estimated standard error of the differences (jointly for the slope and threshold parameters) leading to a chi-square test statistic (Divgi, 1985).

Percentile to theta translation. In computing test accuracy at a particular point, we need to know the theta value corresponding to each point in order to compute expected observed score distributions (using an IRT model) and then classification errors. We examined three ways of linking theta and percentile scores. These were: (1) assume a normal distribution on the latent (theta) scale and use the inverse of the cumulative normal distribution function to map percentiles onto theta; (2) compute the distribution of theta score estimates in the youth population samples and use the inverse of this empirical cumulative distribution function; and (3) sum the posterior theta densities for Youth Population sample examinees and compute a cumulative distribution function based on this composite posterior theta density. Each of the three methods led to very similar results, except at the extremes. The "observed" theta distribution method (method 2) led to the most diverse results at the extremes. We continued with the results from method 3.

Conditional expected observed score distributions. For each percentile point (from 0.5 to 99.5 in increments of 1) we identified the corresponding theta value and used our item parameter estimates to estimate a probability of passing for each item given that theta value. We then used these conditional probabilities to compute the compound binomial distribution giving the probability of each possible number correct score conditional on the underlying theta value (see Lord, 1984). We used operational conversion tables for the Reference Form to convert each number correct score to a percentile score. We thus had estimates of the probability of obtaining each possible *estimated* percentile score for a given *true* percentile score.

Compute classification error rates. Numerical Integration (using the 100 discrete percentile levels) was used to compute expected classification error rates. For each target classification level, we summed the probabilities of a conditional *estimated* (observed) score that was above the classification level across all *true* percentile levels that were *below* than the target to estimate the false positive rate. Similarly, we computed the false negative rate as the likelihood that an examinee will have a *true* percentile level below the target level, but have an *estimated* percentile above the target. We then summed the false positive and false negative rates to get the total classification error rate for each target point on the percentile

scale. (Actually, our computer did most of the summing.) The resulting accuracy profiles for Reference Form WK and GS subtests shown in Figures 1 and 2. The "scallop patterns" that resulted were not fully expected but easy to explain due to the discrete nature of number correct to percentile conversions.

Summary

The results of these illustrative analyses support the feasibility of using expected classification error rates to assess the consequences of different mixes of item difficulty and discrimination levels. If this is so, we will not need to continue with item-by-item "p" value matching. Given initial development of percentile to theta conversions, it takes 10 to 15 minutes to go from a set of item parameter estimates to classification error plots (most of the time is importing and formatting the results in Harvard Graphics), so iterative use of this approach with alternative item sets appears feasible.

The psychometrics is, of course, the easy part. The next step will be to solicit judgments to determine how accurate the forms should be.

References

Feldt, L.S., Staffen, M. & Gupta, N.C. (1985). A comparison of five models for estimating the standard error of measurement at specific score levels. *Applied Psychological Measurement, 9,* 351-361.

Lord, F.M. (1980). *Applications of item response theory to practical testing problems.* Hillsdale, New Jersey: Lawrence Erlbaum.

Lord, F.M. (1984). Standard errors of measurement at different ability levels. *Journal of Educational Measurement, 21,* 239-243.

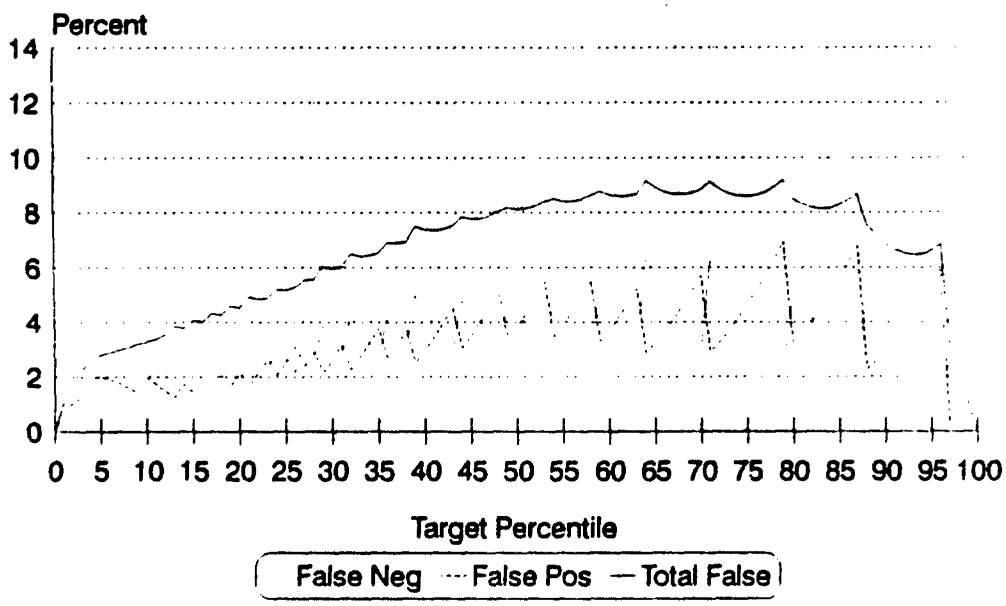
Lord, F. M. & Novick, M. R. (1968). *Statistical theories of mental test scores.* Reading, MA: Addison-Wesley.

Office of the Assistant Secretary of Defense. (1982). *Profile of American youth: 1980 nationwide administrationo f the Armed Services Vocational Aptitude Battery.* Washington, DC: Department of Defense.

Qualls-Payne, A.L. (1992). A comparison of score level estimates of the standard error of measurement. *Journal of Educational Measurement, 29,* 213-225.

Figure 1. Classification Error Rates

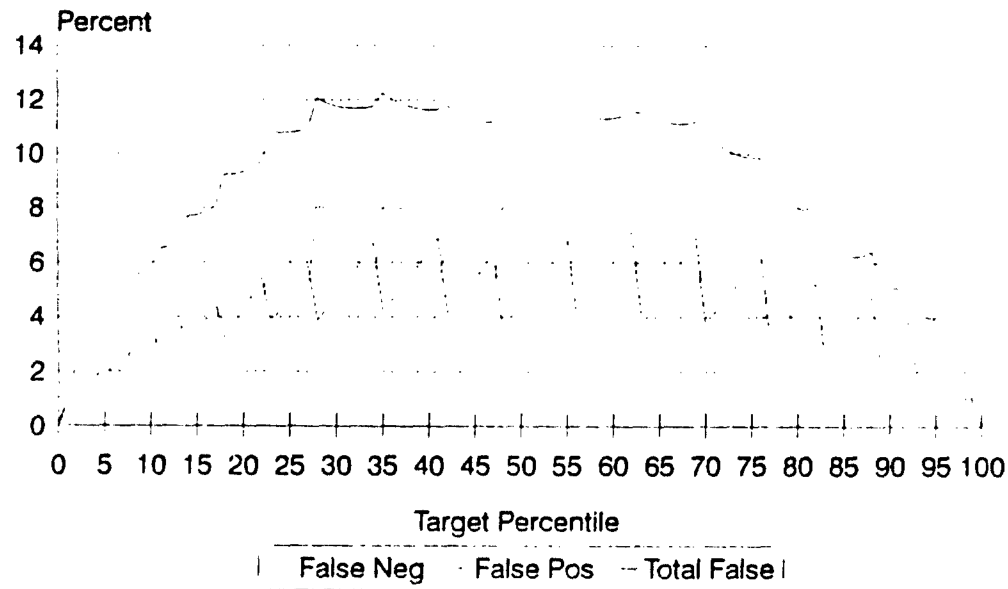
Reference Form WK Subtest



Based on First 2,000 Case Self-Weighted Sample

Figure 2. Classification Error Rates

Reference Form GS Subtest



Based on First 2,000 Case Self-Weighted Sample

THE ASVAB ITEM DEVELOPMENT

Joe Guzaitis and Gretchen Glick
Defense Manpower Data Center

Introduction

In 1989, the Defense Manpower Data Center in Monterey, California, assumed responsibility for the ASVAB, and by fall of 1991, a full complement of psychometricians, test-development editors, and support staff was on hand to begin developing new items for tryout. Supervision of the printing of the current forms 18 and 19 and the operational calibration of the upcoming forms 20, 21, and 22 were also part of the charter of the new group, called the Personnel Testing Division.

The Personnel Testing Division in Monterey consists of eight folks in Test Development and eight in Quality Control and overall management. This paper will detail how the eight of us in Test Development handled the item generation and test-book production for the four subtests that comprise the AFQT portion of the test (Paragraph Comprehension, Word Knowledge, Arithmetic Reasoning, and Mathematics Knowledge)--and then how we are currently handling the development of the so-called "technical" subtests (Auto and Shop Information, Mechanical Comprehension, Electronics Information, and General Science).

Five of us in Test Development deal with the "content" of the items, reviewing them for construct and content validity--asking "Are they good items?". Three others provide support for the development process--coding, archiving, keying, and providing whatever computer expertise we need to produce final camera copy for test book production.

History

The most recent forms of the ASVAB were developed by Operational Technologies Corporation of San Antonio, Texas. Item writing and editing were done by in-house staff, with the exception of some of the AFQT subtests that were typically written by teachers under temporary contract.

Defining a new approach

John Harris, in his paper on the refinement of the ASVAB test specifications, foreshadowed the focus the group was to take toward its approach to ASVAB item development--what can be termed a "closer look." (Harris, 1992.) A closer look at the content domains and test specifications in turn led to a closer look at the methods of item generation and production.

At the item writing and editing level, this meant analysis of previous content with an eye also toward improvement of item timeliness, level of interest, and efficacy. It became the difference between writing a test item and writing a contemporary test item that engages one in the attempting of it and thus contributes to a better metric as well. A tall order, indeed. To be timely, lively, *and* efficacious. To this end we decided to bring writers and subject-matter experts together, create a supportive atmosphere for work, yet closely monitor the actual writing of test items. For the first phase that we started a year ago, that of developing the AFQT subtests, we contracted with BDM, a local consulting agency, for the support materials and facilities to set up a sort of a "city room" atmosphere, reminiscent of an urban newspaper

where writers and researchers work and learn from each other in a feverish atmosphere of shirt-sleeve collegiality.

We placed an advertisement in the local newspaper, inviting writers and math teachers to apply for this full-time, temporary position. Over 78 applications were received: 66 for reading and 12 for math. The number of math applicants was adequate, but for reading it was far beyond our expectations. This reflected the large pool of free lance writers that exists on the Monterey Peninsula, many of them experienced item writers since the peninsula is home to CTB Macmillan/McGraw Hill, the testing arm of the worldwide Macmillan/McGraw Hill publishing firm.

To bring down the large number of reading applicants, we requested a written work sample--a 300-word passage--from those whose resumes showed promise. We provided source material on the high interest topic of "hot air ballooning," and we wrote a sample passage as a criterion. Many of the sample passages surpassed the criterion, and none was significantly deficient. Clearly, we had a motivated, skilled writer pool for the reading subtests. We felt that the more straightforward effort of writing vocabulary items could be used as a reward for those who were high performers in the paragraph comprehension project, which was deemed more difficult. This turned out to be an effective strategy.

We ultimately invited 15 reading and 6 math writers to participate. We had one entire day of item-writing training with the group as a whole. Experienced writers are often surprised to learn of the special skills involved in writing test items. But, experienced writers are also comfortable with manipulating the language, so abbreviated training sessions are often all that is needed. We had lecture sessions on item writing content, style, format, and sensitivity--alternating with writing assignments. These sessions eventually evolved into peer evaluation sessions.

The *Guide to Item Writing* is a 24-page manual we use as both an introduction to test development in training sessions and a reference guide in editorial conferences as item writing takes place. By referring back to a specific that is dealt with in the manual, we are able to provide valuable feedback to a writer that is learning this specialized form of writing.
A second text that is required reading for all item writers (whether new or returning) is our *Bias-Free Testing* guide. This 28-page manual includes guidelines for development of text that provides equal treatment of the sexes and fair representation of minority groups. Further, it details bias-review procedures that include editorial review, peer review, and statistical review.

We introduced the writers to the idea of controlled vocabulary levels for these items. The materials we use are the *Living Word Vocabulary*, a national vocabulary inventory by Dale and O'Rourke, copyright 1976, and *A Revised Core Vocabulary*, copyright 1979 by EDL/McGraw-Hill. A computer software product, PROSE, allows us to look at text as analyzed by several readability formulas; it produces an average grade level number as the readability index.

After introductory training, the two groups--reading and math--broke into separate sections for more specialized training. Each group received the content specifications that John Harris spoke to you about earlier. We had group discussions to make sure writers understood the task at hand--writing of material that would be interesting; engaging; on grade level for content, vocabulary, and skill; and measure the specific skill identified in the specifications.

In developing the paragraph comprehension material, individual writers chose their

passage topics, with repeated caveats from editors against using material which may contain bias against any group or include anything controversial. Writers were also urged to produce passages that dealt with the achievements of minorities and women. The editors kept a topic list in order to ensure diversity; few ideas had to be rejected. This freedom of choice worked well for both writers and editors, to our pleasure.

Passage writers signed out to go to the variety of libraries in the area to do their research. Four community libraries are located nearby to the work site, as well as the Knox library at the Naval Postgraduate School situated in Monterey.

While there were fewer applicants for the math item writers, a small but highly qualified group was assembled. They worked closely together with their own library of source books in what was dubbed the "ivory tower." A communal atmosphere was established, and the group's output, using peer review and close editorial supervision, was also remarkable for the time involved.

A data entry person, who doubled as a receptionist, was continually available to get the raw items and passages in the system. Another editor and I evaluated passages and items and provided feedback on a continual basis. When it was felt that the group could benefit from a joint session in some fine point of item construction, a meeting was called and the point was illustrated and discussed with the group. By the second week, our "item engine" was humming along smoothly. We had figured a reject rate of about 30 percent of the output based on prior item writing experience but found that we were rejecting only about 1 percent. The high selectivity of the applicants, the intense training, and real-time feedback had paid off.

A Control Group

A "control group" concept, to provide a benchmark for our DMDC activities, was suggested early on. A contractor was found, Assessment Systems, Inc. of Minneapolis, Minnesota, to provide a portion of the needed items so that we could determine the efficacy of bringing all item development under the control of DMDC. We provided Assessment Systems with item specifications, format guidelines, and a proposed schedule for completion. Their items received a style edit from us in Monterey, just as we gave to the items developed locally, before the camera copy was produced. Although these items were kept separate during the tryout, the data was compiled in the same format as the other tryout items for easy comparison.

While the initial cost quoted by Assessment Systems was quite competitive to the in-house cost of item preparation, the contractor admitted after the project ended that they had seriously underbid the project and would have to raise their bid substantially for any future projects of this nature. Also, their items arrived late, after we had put together the tryout series A, B, and C; so the control items are in the series D books, the last ones to be administered. To date, we still do not have item-analysis information on those items.

So far, use of a contractor has appeared to provide neither a cost- nor time-effective alternative to in-house item development.

Item Production

After the writing phase had been completed, the items generated were reviewed again and edited several more times for content checks, source checks, technical editing, and copy editing. Final edited items were assembled into tryout subtest files and brought into Ventura

Publisher for desk-top publishing: that is, typesetting and proper page formatting with art in place. A style sheet was developed and used to format the four subtests in similar fashion. Proofs went through first pages, second pages, and camera copy. Answer keys were created with the camera copy and checked by three independent checkers proofreading aloud.

The Personnel Testing Division was able to acquire modem equipment that made for a very efficient work environment:

Publications System interconnected by Novell Netware LAN

Hardware

Unisys 386-20 PCs for writing, editing, and desktop publishing
Mercury MegaDrive removable media for item security
Gateway 486-50 PC for graphic arts
Hewlett-Packard IIIsi high speed Postscript Printer
Kurtzweill Scanner

Software

Windows 3.1
WordPerfect for Windows 5.1
Ventura Publisher for Windows 3.0
CorelDraw 2.0

Camera copy was delivered to the Government Printing Office at Treasure Island. Blue line proofs of the film were then reviewed, as were first press samples of each form. The GPO also assembled, packed, and shipped the tryout booklets to the test sites.

Phase 2

As we approached the next writing phase this summer, that of developing the technical subtests of General Science, Mechanical Comprehension, Electronics, and Auto and Shop Information, we faced a new challenge--finding technical subject matter experts in these specific fields as well as training them to be productive item writers. Working with experienced writers was one thing; this we felt, would be quite another.

Again, a local contractor was engaged, this time the Human Resource Research Organization, HumRRO. They leased and furnished facilities, purchased current trade books and supplies, and coordinated the hiring process while we updated and produced our training materials.

Ads for teachers of science and the industrial arts were placed in several area papers, and notices sent to local high schools and community colleges. We received a total of 113 resumes, sent follow-up questionnaires to 55, and invited 32 to come to an orientation and take a 20-item "ASVAB-like" test to both provide a reality check on whether the applicants knew the rudiments of their specialty and to give them the experience of actually taking some ASVAB technical items to see what they are like and also get a feel for the skill and vocabulary level we would need. Applicants were also given four sample item stems and asked to choose two for completion. We emphasized writing plausible answer choices: one good test of a successful item writer's craft. Those who performed best on this screening exercise (a total of 16) were notified of the time and place of the first training session.

The training sessions were similar to those given for the AFQT portion: a presentation

was made on item writing content, style, and format based on the *Guide to Item Writing* manual; another was made on the topic of sensitivity in publishing based on the *Bias-Free Testing* manual. Subsequent sessions dealt with peer evaluation of draft items.

As training and actual writing began, we found the writers engaged in a variety of approaches to the work. Some subject matter experts became good item writers; some experienced item writers researched the intricacies of the subject matter by interviewing subject matter experts; and some writers and experts worked in a team fashion to mutually explore item writing in a given subject matter area. All three methods worked simultaneously; the staff sorted themselves out, as adults tend to do when left to solve a problem.

As we speak, the item development is ongoing: item writers are responsible for raw items and thumbnail sketches of any artwork. A data entry person takes handwritten or rough-typed originals and puts them into WordPerfect. At this point the editor reviews the word-processed material, along with any thumbnail sketches as necessary. Once the editor and writer agree on an item, a technical illustrator provides finished artwork in CorelDraw from the thumbnail sketches.

You may note that the "control system" is not a factor in this item development phase. Subject matter experts in the Monterey area are producing all the items for the General Science and Technical subtests.

Conclusion

The following methods of item development and production were discussed in light of the current activities of the Personnel Testing Division of the Defense Manpower Data Center.

Historical: Contractor operating under broad guidelines and considerable freedom to recruit SMEs to construct ASVAB subtests conforming to currently acceptable psychometric practice including publication of forms. (OPTech)

Alternate: Item writing contractor utilized for item generation but not for production. This was judged less effective than total in-house form development. (Assessment Systems, Inc.)

Current: Closely managed recruiting and supervision of SMEs and professional writers producing items according to detailed specifications using on- or near-site contractor facilities; intradepartmental form construction and desktop publishing. (DMDC)

We feel that the success of this first effort of ASVAB test form development has been the combined result of

1. the depth of staff test development experience (89 combined years of experience for the 5 principal staff members),

2. the serendipitous location in an area rich in test content experts and item writing personnel available for part-time free-lance activity, and

3. a newly outfitted publishing organization with state-of-the art desktop publishing and graphic arts technology.

On the basis of our first experiences, we are encouraged as to the item-fit-to-specifications as well as the cost effectiveness and overall control of the final product. While this may appear to some to be inappropriate agency micromanagement of a complex development process, we submit that this is the appropriate level of control for a test development effort that has proved to be difficult to manage in the conventional agency/contractor-dependent manner. We are pleased with these first efforts, and recommend this test development approach to any agency.

Reference

Harris, J. A. (1992) The ASVAB test content specifications. *Proceedings of the 33rd Annual Conference of the Military Testing Association.*

THE ASVAB ITEM TRYOUT STUDY

John Welsh
Defense Manpower Data Center

Introduction

One of the changes that was initiated with the inception of the Personnel Testing Division (PTD) was the way in which raw test items were generated for the development of new forms of the battery. The method in which the raw test items are initially tried-out also changed. The items, developed through the use of in-house editors and item writers trained by the PTD editorial staff, are tried out on a broader range sample of recruits in all the Services' recruit reception centers and depots. Forms 2-22 of the Armed Services Vocational Aptitude Battery (ASVAB) were developed under contract and raw items initially tried out on a sample of Air Force recruits - - with target sample sizes of around 500 recruit responses per item. Items surviving this initial round of culling were then candidate for inclusion in over-length forms for the second stage of testing under the old system of building ASVABs.

The old method of trying out raw items, while serving its purpose for over a decade, had a number of draw backs. The first was that items were generated by contractors, usually in a geographically separate part of the country, to written specifications. The raw items generated under this system were then given to ASVAB test developers in booklet form for tryout with relatively few recruits in a highly selected sample (Welsh, Kucinkas & Curran, 1990). Under this development process, a great many items did not survive this initial tryout. Survival rates (of raw items meeting p-value, biserial, and distractor biserial culling criteria) for items of one acceptable item for every three raw items tried out were the rule. Excellent raw item development efforts yielded survival rates of one item for every two raw items developed.

One of the charter objectives of the PTD is to bring the ASVAB development process in-house. This includes use of in-house editors to train subject matter experts in item development - - as you've heard Gretchen Glick describe. Other aspects of this part of the PTD's mission include production of camera copies of all support material, administration manuals, tryout booklets, and answer sheets. It also includes design of tryout studies, data collection including scanning of answer sheets, item analyses and candidate ASVAB form assembly. This part of the symposium will present the design of the tryout of new Armed Forces Qualification Test (AFQT) test items developed by the PTD.

Tryout Study Design

There were many objectives of the tryout study. This feature alone makes this study unique in the development of ASVABs, since past tryout efforts only sought to get item level data and nothing more (Welsh et al, 1990). The primary goal of the study was to tryout over 2200 new ASVAB test items written for the four subtests of the AFQT that would eventually be used in Forms 23 and 24 (slated for use in the DOD Student Testing Program- STP). Over 1700 of these items were written by the subject matter experts hired and trained by the editorial staff of the PTD. Many of the items were written by DMDC staff. The balance of the new items were produced under contract to written specifications sent to the contractor in a geographically distant region of the country. A secondary goal of the tryout study was to compare items developed in-house, through the use of our editorial staff and intensely trained item writers, to items generated in the "old" manner, that is, under contract. Other goals included increased target sample sizes of about 1,000 responses per item in order to obtain

more stable item statistics, sample sizes large enough to compute IRT item statistics, and to test the new items for item position or order effects. All of this was to be accomplished while keeping the testing time in recruit reception centers to under 2.5 hours, testing alternative approaches to IRT parameter estimation based on an equivalent groups design, and testing the equivalence of the groups based on aptitude using recruits' scores of record, adjusting item statistics for observed group differences, and improve the overall survival rates of the newly written items.

The design that would accomplish all of this is presented in Tables 1 and 2.

Table 1. Number of Items

SET A

BOOK	AR		WK		PC		MK		TOTAL
A1	Ref.	30	Ref.	35	Ref.	15	Ref.	25	140
	Fill	8	Fill	5	New	17	Fill	5	
A2	New	38	New	40	New	32	New	30	140
A3	New	38	New	40	New	32	New	30	140
A4	New	38	New	40	New	32	New	30	140
A5	New	38	New	40	New	32	New	30	140
A6	A5	10	A5	10	A5	8	A5	8	
	(2nd qtr.)		(---)		(---)		(---)		
	A2	9	A2	10	A2	8	A2	7	
	(1st qtr.)		(---)		(---)		(---)		
	A3	10	A3	10	A3	8	A3	8	
	(1st qtr.)		(---)		(---)		(---)		
	A4	9	A4	10	A4	8	A4	7	
	(1st qtr.)		(---)		(---)		(---)		
	(Take blocks of items from A2-A5 and arrange in order listed to assess item order effects.)								
App.									140
Test									
Time		46min		18min		26min		30min	120min
Administration time =									30 min
TOTAL approximate time									150 min

Ref. - Reference Group (Normed) items ;Fill - Unscored items - used to equalize test book length; New - Raw items; Rept.- Repeated Items; AR = Arithmetic Reasoning; WK = Word Knowledge; PC = Paragraph Comprehension; MK = Mathematics Knowledge.

As implied by Tables 1 and 2; the design that accomplished all of these goals for us involved the use of four sets (labeled A,B,C,D) of six tryout booklets administered in an equivalent groups design that used administration of the like named reference test in each set of booklets. The reference test items were the only items that were constant across sets (with a few exceptions which I'll discuss shortly). Table 1 indicates that the structure of the A set; but sets B and D are structurally similar. Table 2 shows the composition of the C booklets which were constructed to test for item-order effects and compare the embedded anchor approach to equivalent or random group approach to anchoring the new items.

Table 2. Numbers of Items

SET C

BOOK	AR		WK		PC		MK		TOTAL
C1	Ref.	30	Ref.	35	Ref.	15	Ref.	25	140
	Fill	8	Fill	5	New	17	Fill	5	
C2	New	38	New	40	New	32	New	30	140
C3	New	38	New	40	New	32	New	30	140
C4	New	38	New	40	New	32	New	30	140
C5	New[1]	20	New	20	New	32	New	15	140
	Ref	18	Ref	20			Ref	15	
C6	C5	10	C5	10	C5	8	C5	8	
	(4th qtr.)	(---)		(---)		(---)			
	C2	9	C2	10	C2	8	C2	7	
	(4th qtr.)	(---)		(---)		(---)			
	C3	9	C3	10	C3	8	C3	8	
	(4th qtr.)	(---)		(---)		(---)			
	C4	10	C4	10	C4	8	C4	7	
	(3rd qtr.)	(---)		(---)		(---)			
	Take blocks of items from A2-A5 and arrange in order listed to assess item order effects.)								
App.									140
Test Time	46 min		18 min		26 min		30 min		120 min

Administration time = 30 min

TOTAL approximate time 150 min

Abbreviations same as in Table 1.

The reference test was always the first test booklet in each set. All the booklets containing new items were administered in subtests about 20 -25% longer than operational length subtests. Since the reference forms only exist in operational length, the appropriate number of items was used to fill the reference booklets to be the same length as the over-length, new-item books.. These fill items provide useful data for tryout as well as those in the remainder of the booklets that were tried out with other new items. The testing times were taken from the estimated time-per-item required for the operational subtests - - thus total testing time was estimated from experience with similar item types from the operational battery. All of the sets were constructed so that, not only was the reference test always in the first booklet in the set, but the sixth booklet in the set always contained repeated items from the second through the fifth booklets, rotated in blocks of approximately ¼ subtest length. For example, the block of items from the second quarter of booklet A5 appeared in the first quarter of booklet A6. The second ¼th items in Booklet A6 contained items from the first quarter of booklet A2 and so on as shown in Tables 1 and 2. This feature of the design will allow us to test for systematic differences in the item statistics that could be attributable to the items' position in the test.

The Set C construction depicted in Table 2 has some other unique features. Specifically, items in the C5 booklet were some of the same items as tested in C4, but repeated in corresponding blocks for some of the items in the like-named reference form. This design feature allowed for examination and comparison of differences in item parameters calculated from the equivalent groups design, to item parameters calculated on the basis of embedded item-anchors. This type of information may provide useful information for future tryouts, or for situations where the alternative of using the equivalent groups approach may not prove practical or possible.

Additionally, several analytic goals were incorporated into the design of the tryout study. These included being able to adjust the item parameters of classical statistics to make them comparable to estimates for items tried out at other times. Also, the item analyses will include the use of Differential Item Functioning (DIF) indices for the first time in the development of candidate ASVABs. These indices include the Mantel-Haenszel chi-square and the Educational Testing Service (ETS) delta - - computed from the Mantel-Haenszel Odds Ratio (Holland and Thayer, 1986)

Tryout Results

Figure 1.Biserial Distribtuions for New AFQT Items - Set A

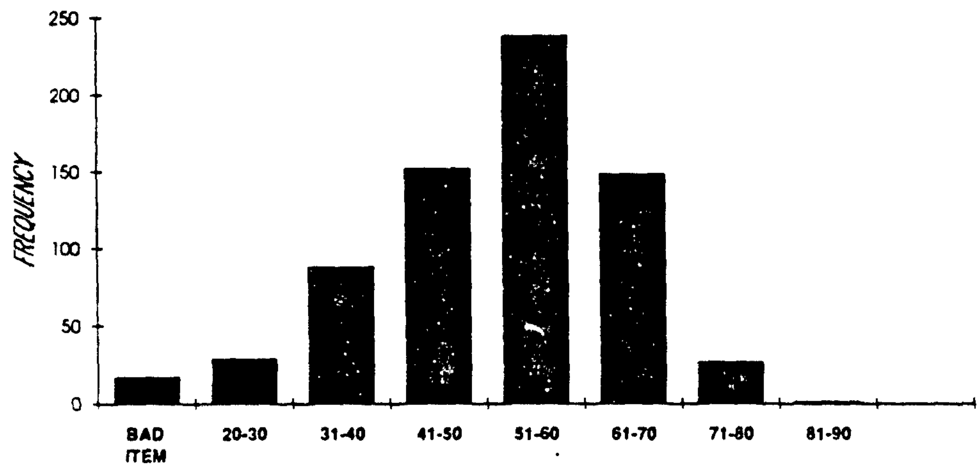
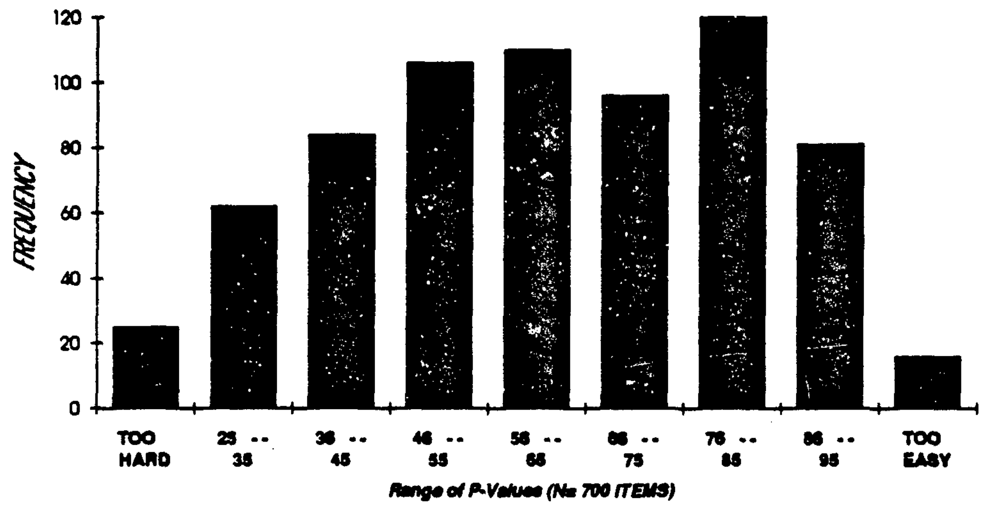


Figure 1 shows the distribution of the biserials for the items tried out for the Set A booklets at nine different recruit depots and reception centers throughout the country and across all four Services. As the interested reader can tell by looking at Figures 1 and 2, we were successful in achieving our objectives that relate to survival rates. The biserial and p-value criteria, were established at rbis $\geq .20$ and the p-values within the range $.25 \leq p. \leq 95$. By these two criteria, approximately 6% to 11% of the items were lost in the A and B sets respectively, because their p-values were out of range (We have only analyzed the A and B sets so far since, as of the writing of this paper, we are still testing as some sites with the C and D Sets). These survival rates for the new items have exceeded our fondest hopes. In examining the data for the new items using other criteria, the results indicate that we don't lose many other items due to positive biserials for distractors, or negative biserials for the keyed answers (usually representing items with two right answers or no correct answer).

Figure 2. Range of P-Values New AFQT Items - Set A



In examining the results of the tryout for negative biserials for the keyed response, and positive biserials for the distractors, we lost 17 items due to negative answers and another 25 items for positive distractors' biserials for a total of about 6% more items lost by these criteria. In total, we lost about 12% to 18% of the items to p-value problems or biserial problems. Of course, these criteria are approximate, since many of the items may be fixed by editing the stems or distractors, and we have no information yet on items that may display magnitudes of DIF that preclude their use in the item bank at this stage of their development.

I mentioned above another concept that is new to the ASVAB development process - - another fundamental change in the process itself. The PTD is not trying out these items for possible inclusion into over-length operational forms -- as has been the case in the past. These newly developed items are being screened into an ASVAB test item bank. The use of an item bank for the form assembly phase of the new ASVAB test development process represents a new tool in the ASVAB development process. The basic change in the goal of the current tryout study now is to produce items for inclusion in an item bank. Target numbers of items are based more on concerns of breadth of taxonomic coverage, as indicated in John Harris' discussion - - the ASVAB content taxonomy has been refined and expanded. The new test development cycle will not, in all likelihood have an over-length phase as in the past. Instead, test forms will be assembled according to the revised subtest specifications and to considerations of test accuracy that Laurie Wise spoke of. The ASVABs being developed now for the STP in 1997 are still unknown to some degree, since the entire ASVAB program may undergo more basic structural changes in the coming year. In all likelihood, the AFQT portion of the ASVAB will remain constant for the next couple of years and these new items, when coupled with the over 7,000 items currently resident in the ASVAB item bank, will provide a solid basis for the construction of future forms.

Conclusion

Based on the preliminary results and the first screening criteria for the first set of new items, Personnel Testing Division has demonstrated its ability to generate new items of sufficient quantity and quality to develop new AFQTs for follow-on forms of the ASVAB. These new items were developed "in-house" and will be compared to other items developed in a

manner similar to past methods, but it safe to conclude that the manner in which the ASVABs of the future will be constructed has already changed, with more, as yet unspecifiable changes on the horizon. The partial results of the first tryout of PTD developed items indicate the tryout study is a resounding success. Next milestones include the development of items for the technical subtests, and for new tests that may be included in the next generation of ASVABs, and then on to assembly of forms for the ASVABs that will be used in 1997.

References

Holland, P.W., & Thayer, D.T. (1986). *Differential item functioning and the Mantel-Haenszel procedure.* (Technical Report No. 86-89). Princeton, NJ: Educational Testing Service.

Welsh, J. R., Kucinkas, S. & Curran, L. T. (1990). *Armed Services Vocational Aptitude Battery (ASVAB): Integrative review of validity studies.* AFHRL-TR-90-22. Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Analysis Levels and Inter-level Linkages: Moderating Change and Intervention Effects

Winston R. Bennett, Jr.
Texas A&M University
and
Armstrong Laboratory Human Resources Directorate

Hendrick W. Ruck
Armstrong Laboratory Human Resources Directorate

Allen I. Huffcutt
Bradley University

Abstract

Military organizations are dynamic and are composed of multiple levels (e.g., individual, workgroup, and organization, among others). These levels and the linkages among the levels have implications for the measurement of outcomes due to organizational interventions or changes. One of the most salient issues facing organizational researchers is related to measuring the effects of interventions at multiple organizational levels. Interventions or changes which focus on a particular organizational level may be perceived as ineffective because they may not lead quantifiable change at subsequent levels. For example, the effect size for intervention or change which occurs at one level will likely be greatest when outcomes are measured at the same level. Further, the magnitude of the effect size decreases as outcomes are measured at subsequent levels. The linkages between organizational levels may actually function as moderators of the effect size of the interventions. Research to date has focused on organizational levels, in and of themselves, with limited acknowledgement of the role that various linkages might play in moderating the ability to detect change at subsequent levels. With changing military structures and constrained fiscal resources, military researchers will be required to justify expensive programs by quantifying outcomes at multiple levels. This paper will discuss issues related to organizational levels and linkages and their relevance for military research. A basic model of organizational levels and linkages will be used to illustrate the potential impact of these issues on the measurement of change. The need for methods and metrics to evaluate outcomes across levels will be highlighted and discussed.

Introduction

Military organizations are highly dynamic organizations. As such, they adopt numerous strategies to maintain and enhance their mission effectiveness and their productivity. These strategies or interventions can include programs for increasing employee motivation, introducing new technology into the workplace, or developing and implementing training programs for employee growth. In many cases, these interventions are not found to be successful in terms of their impact on the organization, its overall effectiveness, and the productivity of the workforce.

The present paper examines issues related to levels of analysis. First, a simple organizational model will be used to discuss levels of analysis. Second, data requirements for addressing levels of analysis will be highlighted. Third, two potential approaches for addressing levels of analysis issues will be highlighted: a multilevel productivity measurement system and a probability-based simulation of organizational linkages to identify resources and manpower constraints related to organizational decision making. Finally, the potential use of both approaches for evaluating organizational interventions will be explored.

Typically, organizations intervene at one level (e.g., the individual or workgroup in a shop) but measure success in terms of effectiveness and outputs at a more aggregate level (e.g., the division, department, or wing). In addition, research evaluating organizational interventions has typically focused on organizational levels, in and of themselves, without an analysis of the role that various linkages might play in moderating the ability to detect change at subsequent levels. Moreover, changes in evaluation criteria tend to decrease as a function of the distance of the chosen criteria used to evaluate interventions from the point of intervention (Goldstein, 1993). In the training literature, more distal criteria are susceptible to organizational variables such as resource availability (Peters & O'Connor, 1980), opportunities to perform (Ford, Quinones, Sego, Speer Sorra, 1992), and increased workplace demands (mobility or sortie generation), among others.

In the absence of cross-level metrics, organizations will be extremely limited in their ability to justify expensive training or technostructural interventions because the positive effects of these interventions cannot be seen by the organization. Given these issues, the need for developing methods and metrics to address cross-level changes in effectiveness is critical. Two basic approaches which may help researchers to more fully explicate and quantify the nature of linkages between levels and potentially to weight and aggregate information across levels will be proposed. One approach is a multi-level productivity measurement and enhancement system, while the second approach is based upon a probability simulation of simple organizational structure, while the other can be seen as

Relevance of Levels of Analysis Issues in the Military Context

The present military personnel and fiscal environment requires that managers select those interventions for which evidence of effectiveness has been demonstrated. In most cases, this effectiveness cannot be adequately determined. Further, it is somewhat paradoxical that interventions designed to increase unit effectiveness or productivity, in fact, are not found to impact organizational effectiveness, readiness, or productivity. Further, it is naive to assume that changes made at an individual or work group level will be manifest in changes at the organizational level. Most military organizations are composed of many intra-organizational and inter-organizational levels. These levels have implications for questions related to the measurement of outcomes related to interventions such as training. For example, at which level to we focus the intervention e.g., individual, workgroup, division or wing)? At which level are outcomes measured (e.g., individual, workgroup, division, or wing)? What is the question to be answered? How should be measure? What is the purpose of our measurement? Who has control over the variables chosen to be measure (e.g., are changes the variables likely to occur with changes in the behaviors of personnel)? Addressing these questions is critical to the measurement of effectiveness and organizational productivity. Figure 1 provides an illustration of a simple organization and identifies some of the linkages and dynamics that may serve as moderators.

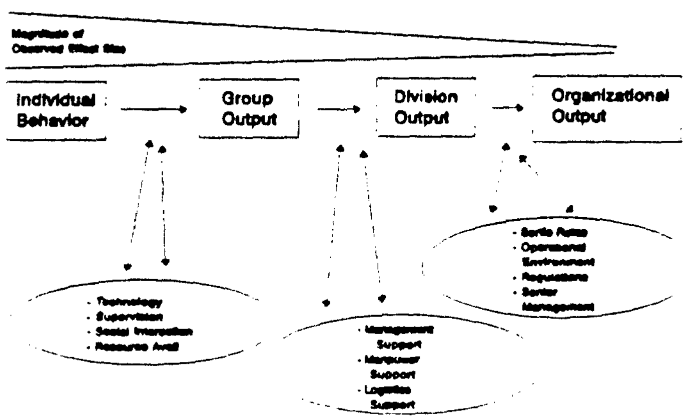


Figure 1. Simple Organizational Structure

Organizations must be sensitive to the fact that there are moderating variables embedded within the organizational structure that influence the ability to detect change due to interventions beyond the level at which the intervention occurred. Attempting to measure workgroup level change at the organizational level, without an understanding of the interplay of linkages and levels of analysis, will result in less than adequate evidence of the benefits of the intervention. Another key point is that organizations can only use productivity measures as evaluation criteria if these are related to the intervention. The intervention must impact indicators that are addressed by the measurement system in order for change to be detected.

Interventions and changes made within levels in the organization, are perceived as ineffective because they do not lead to overall organizational productivity change. This lack of cross-level transferability, and the grain size of the analytic methods employed, reduces the capability of military managers to accurately evaluate the impact of interventions such as training and work redesign. Initially, data required to inform multilevel evaluations must be developed. Subsequently these data can be used within two potential approaches Two potential approaches for addressing levels of analysis. These approaches are (a) the use of multilevel productivity measurement systems to address levels of analysis, and (b) a probability-based simulation approach for modeling organizational change and important linkages will be discussed.

Data Requirements

To adequately address levels of analysis and linkage issues, data must capture multiple organizational levels. That is, the data used to evaluate interventions but be identifiable at different levels. Task level information must be combined into more aggregate information such as that associated with jobs. The job-level information can be further aggregated into division or directorate-level information, and ultimately into higher-order categories. In addition, this aggregation will probably not occur in direct correspondence. At some level there is the likelihood that the same information on tasks will have a contribution to more that one job, or group of jobs, and so on. The amount of overlap of information across aggregates must be quantifiable if the data are to be useful for addressing linkages and levels of analysis issues. Further, as task or individual level information is combined, there will be a loss in the "observed" impact of changes at the lower levels. What is critical, is that there is a quantitative linkage among the aggregated data. This quantitative link will ensure that whatever change is accomplished at the smallest grain-size of analysis (e.g, task, individual, or workgroup) can be tracked to and from the more aggregated levels of information.

In addition, data which provide an indication of the probability of observing change at subsequent level given the potential moderators of that change must be developed. As can be seen in Figure 1, there are numerous potential moderators at each level. The impact of these moderators, among others, needs to be clearly specified. This specification may be accomplished using teams of workers, organizational managers, and researchers to develop rational estimates. What should be possible is the specification of maximum likelihood estimates of the observed effect size of the change at each subsequent level. It should be noted that the point at which further change that can be directly related back to the finer levels of specification cannot be seen, will likely be different for different interventions. However, change should be detectable at levels beyond the level at which the intervention occurred.

Multi-Level Productivity Measurement and Levels of Analysis

Productivity measurement can play a key role as an intervention for monitoring work activities and providing feedback related to effectiveness and as a criterion for other interventions such as training and technostructural change. A productivity measurement approach which may provide information to be used to address organizational levels is the Productivity Measurement and Enhancement Systems (ProMES) (Pritchard, 1990).

The development of a productivity approach includes discussions with, and input from, individuals at different levels within organizations (Pritchard, 1992). Typically, indicators of productivity are developed for a target group of individuals (usually at the workgroup level). Typically, measures of each indicator, taken at the same level as that upon which the indicators were developed, are sensitive to changes in workgroup effort. However, since indicators of productivity at subsequent levels are rarely developed, it is difficult to demonstrate

change at more distal organizational levels. Figure 2 provides a simple illustration of the linkage among indicators at different levels.

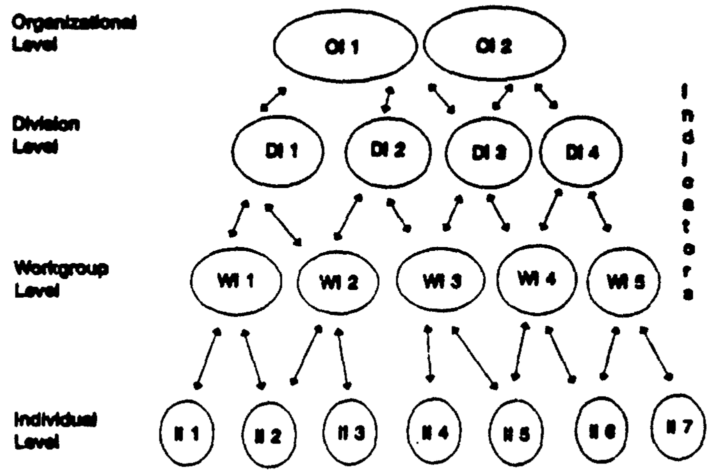


Figure 2. Productivity Measurement Across Levels

In a multi-level productivity approach, development would begin at the "lowest" level in the organization, the individual level or workgroup level. The goal at this level is to identify the products and indicators that reflect those activities for which the individuals or the workgroup have control. In addition, products and indicators for activities which are beyond the control of the individual or work group would be also be identified. The idea is each group of indicators is causally related. Final sets of these indicators would be used by the next level during their product identification and indicator development. Again, there would be products and indicators for which the division has control and those for which it does not. What should happen is that some percentage of the output from the individual or workgroup level will be incorporated into the division level indicators. Eventually, the organization will have a set of products and indicators as well. Contingencies can be used to provide information about the relative weights of the outputs, which in turn might be used to weight change due to an intervention within and across levels. While the impact of individual or workgroup change may not be completely realized at the organizational level, some percent of maximum possible effectiveness can be traced through the organizational levels.

A Probability-Based Simulation of Organizational Levels

The second approach can be illustrated by returning to Figure 1. In the figure, a basic set of linkages among the levels in an organization are highlighted. In addition, a preliminary set of moderating variables are identified for each successive linkage. If an intervention or change is made at the individual level, it is illustrated that the magnitude of the effect size associated with change is greatest at the same level. Figure 1 also illustrates that the magnitude of the effect size decreases as it crosses successive levels. How does this occur?

One answer is that each level and each linkage serves as a separate and significant moderator of the ability to detect an effect. Moreover, there may actually be a quantitative reduction in the effect size due to each of these moderators. These quantitative reductions should be quantifiable, possibly using a probability-based simulation. Research addressing these linkages might attempt to quantify the contribution of changes in productivity at one level and at a second but proximal level. A series of propositions might be addressed by a simulation. Initially, the simulation would be developed by having an interdisciplinary panel of organizational experts develop a general

simulation model of the interplay of the linkages. The linkages in the simulation would be based upon a consensus of the experts related to probabilities associated with the transitions from one level to another in the simulated organization. Probabilities associated with the contribution of various groups of interrelated inputs and outputs at each levels would also be included. Subsequently, the impact of change in the simulation would be tested by conducting sensitivity analyses at various levels in the model and examining outcomes. Basically, hypotheses related to output that might be resultant from changes in several inputs are generated. Inputs at one level are changed, and the overall impact on the model outcomes, in terms of outputs at another level, would be evaluated against a steady state model solution. While there would not be 100% congruence between change at one level and the other, it should be possible to identify and quantify *some* amount of change which is a portion of the overall or maximum effectiveness of the model. Ultimately, the utility of the simulation is as a testbed to explore the quantitative relationship among levels. Outcomes from the simulation will help to enhance the understanding of the potential effect that linkages and levels of analysis upon organizational change.

Conclusions

In the future, civilian and military researchers will be faced with a greater need to justify their intervention programs on both qualitative and quantitative terms. This justification will necessarily include assessments of the impact of intervention programs at multiple levels. Given that these levels potentially moderate the effectiveness of intervention programs, it is prudent to address the impact of the moderating influences upon the observed effectiveness of the program. Understanding the nature of the impact of levels on outcomes will be central to the continued viability of training programs and other interventions. As the military manpower and training budgets are reduced, researchers and managers will need quantitative evidence of the usefulness of their programs.

The present paper highlighted several of the problems associated with levels of analysis and linkages among organizational levels. In addition, two approaches for describing the potential impact of crossing levels in evaluation and for the development of quantitative indicators of the role of levels of analysis and linkages were proposed and described. Both approaches have promise for testing many of the issues highlighted in this paper. The military environment offers a rich testbed for the development and exploration of methods related these issues. Future research should begin to explore the empirical issues related to linkages and levels of analysis within the military context.

References

Ford, J. K., Quinones, M. A., Sego, D. J., & Speer Sorra, J. (1992). Factors affecting the opportunity to perform trained tasks on the job. *Personnel Psychology, 45*, 511-527.

Goldstein, I. (1993). *Training in organizations* (3rd ed.). Pacific Grove: Brooks/Cole.

Peters, L. H., & O'Connor, E. J. (1980). Situational constraints and work outcomes: The influence of a frequently overlooked construct. *Academy of Management Review, 5*, 391-397.

Pritchard, R. D. (1992, April). *Issues in productivity measurement.* Seminar in productivity conducted at Texas A&M University.

Pritchard, R. D. (1990). *Measuring and improving organizational productivity: A practical guide.* New York: Praeger.

Multilevel Occupational Analysis: Hierarchies of Tasks, Modules, Jobs, and Specialties

Jimmy L. Mitchell
McDonnell Douglas Corporation

William J. Phalen
Armstrong Laboratory Human Resources Directorate

Darryl K. Hand
Metrica Incorporated

Abstract

There are extensive amounts of task-level information now available on most military occupations which are used effectively for a variety of purposes. Organizing such task information into task modules, jobs, and higher order categories allows the data to be applied to more global issues and problems; the summarized data can be used to develop more realistic models or simulations of occupational structures and requirements. Existing data already permit comprehensive organizational modeling; some present analyses involve multiple specialties, multiple categories of personnel (enlisted, officer, civilian), or even multiple services (interservice or joint service projects). Given the substantial value of task-based information and analyses, multi-level studies focused on task modules and other higher order groupings have considerable potential for applications in modeling military organizations to assist military decision makers in evaluating proposed organizational restructuring, interventions, and/or other organizational changes. New ASCII Comprehensive Occupational Data Analysis Programs (CODAP) technology permits analysis of occupational data at a number of different levels of specificity or from a variety of viewpoints. Implications of multi-level data for simulating organizational change will be discussed.

Introduction

The principal occupational analysis technology in the United States Air Force is the Task Inventory/Comprehensive Occupational Data Analysis Programs (CODAP) approach. This system has supported a major occupational research program within the Human Resources Directorate of Armstrong Laboratory (formerly the Air Force Human Resources Laboratory) since 1962 (Christal, 1974), and an operational occupational analysis capability within Air Training Command's USAF Occupational Measurement Squadron (formerly Center) since 1967 (Weissmuller, Tartell, & Phalen, 1988). CODAP is now used by all the U.S. and many allied military services, as well as other government agencies, academic institutions, and some private industries (Christal & Weissmuller, 1988; Mitchell 1988).

Recently, several major new programs were created to extend the capabilities of the system for assisting analysts in identifying and interpreting potentially significant jobs (groups of cases having similar jobs) and task modules (groups of co-performed tasks). Initial operational tests of these automated analysis programs were conducted and results were reported at previous MTA conferences and Occupational Analyst Workshops (Mitchell, Phalen, Haynes, & Hand, 1989; Phalen, Mitchell, & Hand, 1990; Mitchell, Hand, & Phalen, 1991; Mitchell, Phalen, & Hand, 1991). As of the end of July 1992, all of these formerly experimental programs have been transitioned into the operational ASCII CODAP system, and are now available on the USAF Occupational Measurement Squadron's IBM computers as well as on the Armstrong Laboratory UNISYS. CODAP on-line program documentation has also been expanded to include the new analysis assistance programs. On-going operational testing and evaluation of the new interpretive software continues to demonstrate the value of these programs in terms of enhanced analytic capabilities and potential for accelerating completion of an occupational analysis. Current experimental work is focusing on adjusting the task clustering algorithm or expanding the task co-performance similarity matrix to yield more interpretable groupings of tasks, so as to distinuish meaningful subgroups among the large numbers of commonly performed tasks.

* Approved for Public Release: Export Authority 22CFR 125.4 (b)(13)

Multiple Levels of Analysis

While detailed task-level information is critical for certain uses, such as training development or selecting topics for promotion tests, such data may be too complex and specific for other purposes; for example, for facilitating management macro-level decision making or evaluation of possible impacts of organizational restructuring. Organizing task information into task modules, jobs, and higher order categories allows the data to be applied to more global issues and problems and can be used to develop realistic models or simulations of occupational structures and requirements.

Task clustering using co-performance values as a basis for developing task modules (TMs) has been reported elsewhere and need not be detailed again here (Perrin, et al, 1988; Vaughan, et al, 1989; Mitchell, Phalen, & Hand, 1991). Task co-performance is defined as a measure of the similarity of pairs of task profiles across all the people in an occupational survey sample. For details of the computation of measures of task co-performance, see Rue, Rogers, and Phalen (1992).

Task Module-Level Job Descriptions

TM-level data can be used to provide very concise descriptions with which to compare jobs within a specialty (see Figure 1). Note that time spent and percent performing data are average values across the tasks within each module; this provides comparable statistics for the TMs. It is much easier to compare such TM-level job descriptions for various jobs within a specialty than it is to analyze task-level job descriptions (typically ordered on decending percent time spent or percent performing. Thus, the TMs provide a structured, summarized set of data which facilitates between or among job comparisons. In this particular display, the display is ordered on decending average percent time spent; this brings the most "important" (in terms of average job time per task) areas of work to the top of the job description to quickly highlight job specialization. Note that if the Sum of time spent were used to order the display, it is the more general task modules which would head the list. While that is, of course, an accurate description, such a display tends to emphasize commonly-performed tasks and to disguise the specialization or uniqueness of a job group.

PRTMOD Mobile Systems Maintenance (n = 157) **Page 2**

TM	Module Title	No. of Tasks	Percent Time Spent Sum	Percent Time Spent Cum	Percent Time Spent Avg	Average Percent Members Performing
31	Maintain Moble Computers or Switching Systems	1	.55	.55	.55	69.43
01	Generic Maintenance Tasks	19	8.57	9.12	.45	71.81
02	Perform General Maintenance	30	12.76	21 88	.43	70.96
06	Maintain Magnetic Tape Units	7	2.96	24.84	.42	68.70
20	Grounding Systems, Cables, & Wiring	6	2.26	27.10	.38	70.70
07	Console and Operator Panel Maintenance	8	2.98	30.08	.37	68.23
09	Power Supply Maintenance	4	1.32	31.40	.33	64.97
08	Processor & Memory Assemblies Maintenance	15	4.91	36.31	.33	60.81
05	Maintain Display Equipment	6	1.93	38.24	.32	61.68
18	Power Distribution System Maintenance	5	1.53	39.77	.31	61.02
84	Supervise AFS 30554 Personnel	1	.26	40.03	.26	47.13
16	Maintain Modems	9	2.29	42.32	.25	52.02
04	Maintain Printers	4	.93	43.25	.23	50.16
19	Maintain Batteries & Battery Chargers	9	2.02	45.27	.22	43.45
51	Deployment Preparations	11	2.47	47.74	.22	35.55
27	Check and Repair Telephones and Other Sets	15	3.16	50.90	.21	37.96
28	Troubleshoot Fixed or Mobile Trunk Circuits	3	.60	51.50	.20	33.97
.	.	.	.	.	.	.
.	.	.	.	.	.	.

Figure 1. Example Task Module-Level Job Description - Mobile Systems Maintenance (AFS 305X4)

If more detail is needed, data for individual tasks can be displayed within each TM (see Figure 2); this gives a perspective on how TM-level data are derived and structures task data into a more easily comprehended structure.

PRIMOD Mobile Systems Maintenance (n = 157)					Page 2
TM Module Title	No. of Tasks	Percent Time Spent Sum	Cum	Avg	Average Percent Members Performing
31 MAINTAIN MOBLE COMPUTERS OR SWITCHING SYSTEMS	1	.55	.55	.55	69.43
L 623 Perform PMIs on mobile electronic computer or switching equipment		.55	.55	.55	69.43
01 GENERIC MAINTENANCE TASKS	19	8.57	9.12	.45	71.81
F 325 Perform general cleanging of electronic computer or switching systems equipment		.70	.70		92.99
K 618 Replace minor electrical hardware, such as lamps, switches, fuses, or connectors		.70	1.41		90.45
K 619 Replace nonelectrical hardware, such as screws, nuts, ejectors, or covers		.62	2.03		89.17
F 326 Perform inspections of cables, cable troughs, or connectors		.60	2.63		90.45
F 315 Locate units, modules, rows, columns, components, pins, connectors, or test points using alpha/numbric designators		.59	3.22		84.08
F 373 Solder or desolder electronic equipment components		.52	3.74		82.80
F 308 Connect or disconnect power or equipment leads		.48	4.21		82.80
F 327 Perform inspections of nonelectrical hardware		.47	4.68		78.98
F 328 Perform inspections of power or grounding systems		.46	5.14		82.80
E 275 Research microfiche documents for parts information		.45	5.59		71.97
F 306 Clean or treat filters		.43	6.03		73.89
E 276 Research publications for parts numbers		.41	6.43		66.24
K 557 Remove or replace batteries		.40	6.83		74.52
F 323 Perform corrosion control, other than general cleaning		.38	7.21		64.97
H 415 Perform PMIs on systems, such as cabinets, racks, or subfloors		.36	7.57		57.96
K 575 Remove or replace filters		.33	7.89		61.15
F 307 Comply with TCTOs		.31	8.21		61.78
E 258 Maintain or issue consolidated tool kits (CTK)		.21	8.42		33.76
E 255 Maintain equipment operation logs		.15	8.57		23.57
02 PERFORM GENERAL MAINTENANCE	30	12.76	21.88	.43	70.96
F 366 Perform power-up or power-down procedures		.76	.76		96.18
J 447 Interpret block diagrams for fault isolation		.63	1.39		95.54
J 446 Discriminate between hardware and software failures		.61	2.00		89.81
J 451 Interpret results of diagnostic programs		.58	2.58		85.99
J 448 Interpret logic diagrams for fault isolation		.58	3.15		89.81
J 452 Interpret schematic diagrams for fault isolation		.55	3.71		85.99
. . .	.	.	.	.	.

Figure 2. Example Task-Level Data Within Module Job Description - Mobile Systems Maintenance (AFS 305X4)

This kind of information can be used in a variety of ways, to explicitly define each module when TM-level ratings are gathered or for use in the development of detailed OJT for a particular job. Once SMEs are exposed to such task-based descriptions of the TMs, they can then use the TMs alone (as in Figure 1) as a basis for defining possible changes in the jobs or training programs of a specialty. In those situations where several closely related specialties are being considered together (as, for example, in studies of possible AFS mergers), TM-level information may serve as an adequate basis for making job or training program restructuring decisions.

Evaluating Varying Levels of Specificity

In the current Training Decisions System R&D effort, we are also examining the potential impact of varying levels of specificity of task modules. Automated clustering and initial TM validation at the technical training

center yielded some very large modules; in the Aerospace Propulsion (AFS 454X0) study, TM four contained 82 tasks involving Engine inspection and repair. Based on task co-performance, if an individual performs some of these tasks, he or she is very likely to perform most or all of the tasks. However, for TDS purposes, we collect allocation (learning curves) data on the hours of training (all types) needed for the average Aerospace Propulsion specialist to become proficient on these tasks (that is, able to work independently with minimal supervision). These data are used to quantify the training requirements for the TM (see Mitchell & Lamb, 1989).

SMEs at operational bases criticized the number of tasks in three TMs, and suggested dividing TM into smaller, more internally consistent TMs (see Figure 3). The SMEs stated that getting ratings of time spent in various training settings for the very large modules was unrealistic; they suggested that it would be much more reasonable and accurate to rater the set of smaller TMs. This is, of course, a good research question: does it make a difference? Can you get more reliable, valid ratings using the smaller groupings of tasks? Does this added complexity (more TMs and thus more ratings) result in better estimates (OJT costs, OJT capacity, etc.) or are the less complex, more global modules adequate for the TDS level of macro-analysis? We are currently evaluating these questions; we have collected allocation estimates using both sets of TMs and, once the TDS data base for this AFS is complete, will be using both the sets in running problems and making estimations. A preliminary look at the ratings collected suggests there are some differences (with the aggregate total of the more discrete TMs exceeding the more global estimates), but we do not yet know whether these differences will substantially impact model outputs. We should have an answer to this question within the next few months.

This type of subclustering also raises the possibility that we might want more, smaller, more discrete TMs for some purposes, while using fewer, larger TMs to meet other needs. For specific on-the-job training management, a local trainer might want the more discrete TMs to guide the development of an OJT program. A functional manager, however, might prefer to use the larger, more generic TMs when making macro-level decisions about formal training programs, when comparing multiple specialties, or when communicating the needs of the specialty to higher level managers, manpower staff, or personnel policy analysts. Thus, the level of specificity of task clustering may depend, in part, on the specific use that is to be made of the data. One way to accomodate both extremes is to organize the data in hierarchical modules, as displayed in Figure 3. This provides the flexibility whereby the particular user may chose the level of detail he or she wishes for the particular purpose at hand.

Refinements in Automated Task Clustering

If field SMEs find very large TMs unacceptable, how can we adjust the automated task clustering algorithms so as to produce smaller, more internally consistent modules? In a further ASCII CODAP effort, an attempt was made to weight the similarity values on which the task clustering is based so as to penalize very large groupings of tasks. Experimental runs for four AFSs were generated with the new weighted algorithm, but the results were not particularly encouraging. The resulting TMs were about the same so that this approach does not appear to have any significant effect. Likewise, this change in algorithm appeared to have no impact on TMs where very few people perform. Clearly, another approach is needed for further refinement of the automated task cluster analysis programs. Since the task clustering algorithms work well for most AFSs and TMs, the program was operationalized; that is the programs were moved from experimental status to be available to any CODAP user on the UNISYS. In August 1992, these new programs were transported to the USAFOMSq IBM equipment so as to be generally available, and program documentation revised (as of 31 July) to explain their use.

A further research approach was also devised to explore other refinements to task clustering, by expanding the similarity matrix to include equipment or system operated or maintained, or other background data. We may also explore the possibility of using TD and TE data in some way to help further refine the modules; this would clearly be in line with the expectation that tasks within a TM should share common skills and knowledges.

Conclusions

The work on the MODULES programs to date has been highly successful and we have only begun to tap the potential of this type of automated modular technology. Further, our work with TMs to date has led us to believe that the whole module approach has real promise for simplifying and expanding the use of occupational information in helping executives and managers make more realistic decisions; this is a critical technology in the current period of manpower and budget reductions and consolidations. It is more important than ever to be able to model proposed changes and assess the potential impact of such changes before final decisions are made.

Task Module 0004 - Engine Inspection and Repair

4 a. General Cleaning and Blending

G 249 Blend engine compressor blades
G 250 Blend engine fan blades
G 251 Blend engine inlet vanes
G 252 Blend engine stator vanes
G 253 Blend engine turbine air seals
G 254 Blend engine turbine nozzle vanes
G 255 Blend engine turbine wheel blades
G 256 Clean engine parts using cleaners, other than ultrasonic cleaners
G 257 Clean engines
G 262 Drain fuel filters

4 b. General Inspection

G 265 Inspect accessory gearboxes
G 268 Inspect engine bleed valves and actuators
G 269 Inspect engine combustion sections
G 270 Inspect engine compressors
G 271 Inspect engine controls
G 272 Inspect engine electrical components
G 273 Inspect engine exhaust section components
G 274 Inspect engine fan section components
G 275 Inspect engine hydraulic systems
G 276 Inspect engine oil filters
G 277 Inspect engine or accessory splines
G 278 Inspect engine plumbing
G 279 Inspect engine stator vanes
G 282 Inspect engines before or after operation
G 283 Inspect fuel filters
G 285 Inspect gearbox assemblies
G 287 Inspect inlet guide vane (IGV) actuating systems
G 288 Inspect magnetic engine chip detectors
G 290 Inspect nozzle position transmitters
G 297 Inspect turbine exhaust cases
G 298 Inspect turbine nozzles
G 299 Inspect turbine rotor blades
G 300 Inspect turbine rotors
G 301 Inspect turbine unit assemblies
G 303 Inspect variable stator vanes (VSV)

4 c. General service and repairs

G 358 Perform flex boroscopeinspections of engines
G 370 Perform rigid boroscope inspections of engines
G 377 Pressure check engines prior to operation
G 420 Remove or install IGV actuating system components
G 423 Remove or install magnetic engine chip detectors
G 451 Repair engine fuel nozzles
G 457 Rig IGV systems
G 467 Service CSD systems In Shop Inspection
K 615 Inspect engine exhaust plug nuts
K 616 Inspect engine fuel manifolds
K 617 Inspect engine fuel nozzles
K 620 Inspect engines removed from storage
K 621 Inspect fuel manifold test stands
K 624 Inspect secondary thrust reverser nozzles
K 625 Inspect ultrasonic cleaners
K 628 Lap engine oil carbon seals

4 d. Maintain Shop equipment

K 630 Maintain bearing servicing equipment
K 631 Maintain engine accessory shop equipment
K 632 Maintain fuel manifold test stands
K 633 Maintain ultrasonic cleaners

4 e. In Shop Repair and Check-out

H 511 Remove or install dowel pins or drive pins
K 596 Assemble or disassemble cowl latches
K 597 Assemble or disassemble engine actuators
K 598 Assemble or disassemble engine bleed valves
K 601 Bench check engine actuators
K 603 Bench check engine bleed valves
K 605 Bench check or pressure check engine carbon seals
K 637 Perform operational checks of fuel manifolds
K 638 Perform operational checks of fuel nozzles
K 640 Perform wax checks on compressor rotor casings
K 667 Repair engine accessories or components
K 668 Repair engine combustion sections
K 669 Repair engine compressors
K 670 Repair engine fuel manifolds
K 671 Repair engine gearboxes
K 672 Repair engine plumbing
K 673 Repair engine stator vanes
K 674 Repair turbine sections
K 676 Vacuum check engine carbon seals

4 f. In Shop Cleaning and Blade Maintenance

K 606 Blend engine turbine blades
K 607 Clean and inspect engine bearings
K 608 Clean and inspect engine oil seals
K 609 Clean engine parts using ultrasonic cleaners
K 611 Grind webs of compressor wheels
K 612 Grind webs of turbine rotors
K 677 Weigh engine compressor blades
K 678 Weigh engine fan blades

Figure 3. Example of Aerospace Propulsion (AFS 454X0) Complex Task Module.

We believe that the MODULES technology can be extremely useful in modeling occupations and the world of work. With additional refinement, we should be able to use quite a variety of information to develop better TMs which will take into account the type of equipment operated or maintained as well as subjective assessments

such as TE and TD. While not yet fully explored or validated, this emerging TM development methodology has great promise for significant improvement of military manpower, personnel and training planning and decision making, and indeed, for organizational analysis as well.

References

Christal, R.E. (1974). *The United States Air Force occupational research project* (AFHRL-TR-73-75, AD-774 574). Lackland AFB, TX: Occupational Research Division, Air Force Human Resources Laboratory.

Christal, R.E., & Weissmuller, J.J. (1988). Job-task inventory analysis. In S. Gael (Ed), *Job analysis handbook for business, industry, and government*. New York: John Wiley and Sons, Inc. (Chapter 9.3).

Mitchell, J.L. (1988). History of job analysis in military organizations. In S. Gael (Ed), *Job analysis handbook for business, industry, and government*. New York: John Wiley and Sons, Inc. (Chapter 1.3).

Mitchell, J.L., Hand, D.K., & Phalen, W.J. (1991). ASCII CODAP: Modification of the CORCAS program to facilitate analysis of task clusters. *Proceedings of the Seventh International Occupational Analysts Workshop*. San Antonio, TX: USAF Occupational Measurement Squadron.

Mitchell, J.L., & Lamb, T.A. (1989). Quantification of specialty training requirements. In the symposium, The Air Force Training Decisions System: Modeling Management and Policy, H.W. Ruck, chair. *Proceedings of the 31st Annual Conference of the Military Testing Association*. San Antonio, TX: Air Force Human Resources Laboratory and the USAF Occupational Measurement Center.

Mitchell, J.L., Phalen, W.J., & Hand, D.K. (1991, October). ASCII CODAP: Use of CASSET to facilitate analysis of case clusters. *Proceedings of the 33d Annual Conference of the Military Testing Association*. San Antonio, TX: Human Resources Directorate, Armstrong Laboratory and the USAF Occupational Measurement Squadron.

Mitchell, J.L., Phalen, W.J., Haynes, W.R., & Hand, D.K. (1989). *Operational testing of ASCII CODAP job and task clustering methodologies* (AFHRL-TP-88-74; AD-214 727). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Perrin, B.M., Knight, J.R., Mitchell, J.L., Vaughan, D.S., & Yadrick, R.M. (1988, September). *Training decisions system: Development of the task characteristics subsystem* (AFHRL-TR-88-15, AD-A199 094). Brooks AFB, TX: Training Systems Division, Air Force Human Resources Laboratory.

Phalen, W.J., Mitchell, J.L., & Hand, D.K. (1990). ASCII CODAP: Progress report on applications of advanced occupational analysis software. *Proceedings of the 32nd Annual Conference of the Military Testing Association*. Orange Beach, FL: Naval Education & Training Program Management Support Activity.

Rue, R.C., Rogers, S.G., & Phalen, W.J. (1992, May). Comprehensive occupational data analysis programs (CODAP): Computation of measures of task co-performance. Manuscript report; SRA Corporation and AL/HRM, available from the authors. Brooks AFB, TX: Armstrong Laboratory, Human Resources Directorate, Manpower and Personnel Research Division.

Vaughan, D.S., Mitchell, J.L., Yadrick, R.M., Perrin, B.M., Knight, J.R., Eschenbrenner, A.J., Rueter, F.H., & Feldsott, S. (1989, June). *Research and Development of the Training Decisions System* (AFHRL-TR-88-50). Brooks AFB, TX: Training Systems Division, Air Force Human Resources Laboratory.

Weissmuller, J.J., Tartell, J.E., & Phalen, W.J. (1988). Introduction to operational ASCII CODAP: An overview. *Proceedings of the 30th Annual Conference of the Military Testing Association*. Arlington, VA: U.S.Army Research Institute.

An Organizational Analysis Simulation Technology

David S. Vaughan
McDonnell Douglas Corporation

Robert M. Yadrick
Armstrong Laboratory Human Resources Directorate

Abstract

One approach to addressing the quantifiable reduction in effect size which is attributable to the moderating effect of linkages is to develop a model which quantifies the contribution of changes in performance and/or productivity at one level and at other proximal levels. Such a model would express important macro-level outcome variables (e.g., organizational productivity, mission effectiveness) as functions of important causal variables. These include both micro-level variables that are directly manipulated in organizational interventions and more macro-level variables, such as business and economic conditions. This type of model would be very useful for exploring the quantitative relationships among events at various levels of abstraction. Such a model could be used to determine the maximum impact that micro-level organizational interventions can reasonably be expected to have on macro-level outcome variables, relative to other uncontrolled events. This paper describes the Training Decisions System (TDS), a simulation technology which relates events at various levels of abstraction. The model relates micro-level personnel events to macro-level outcome variables. Data for the TDS comes from a variety of sources, including job analysis, existing manpower, personnel, and training data bases, and subject-matter experts' judgments. The TDS estimates overall job flows, training requirements, and training costs from individual job, task, and training assignments, based on job analysis data. The TDS model first simulates the flow of individuals through jobs, with task performance requirements, and formal training courses. From these individual events, the model estimates task-level, on-the-job training events. Finally, the model estimates overall training resource requirements, costs, and capacities from the task-level events.

Introduction

We, as behavioral scientists, believe that organizational interventions can produce significant improvements in organizational productivity. However, it has proven very difficult to demonstrate that such interventions actually produce improvements in organizational performance indices. These performance indices are typically macro-level measures--they reflect overall organization performance. In contrast, organizational interventions of the types considered here are micro-level--they focus on particular individuals within an organization. Of course, one possible reason for the difficulty in demonstrating organizational performance impacts of interventions is that they really have little influence on organizational performance. However, we believe the more likely reason is that many variables, operating at various organizational levels, impact on overall organizational performance measures. Such variables might include other (controlled or uncontrolled) organizational changes. Such variables also operate outside the scope of the organization; examples of this are technology changes, actions of other organizations, and the overall economic and political climate. In studies of interventions in real organizations, all of these "extraneous" variables, operating at various organizational levels, also effect the outcome measures in uncontrolled ways. Thus, changes due to intervention may be swamped by changes due to these other uncontrolled factors.

* Approved for Public Release: Export Authority 22CFR 125.4(b)(13)

In light of all this, how might we estimate the impact of an organizational intervention on organizational productivity? One approach involves building a formal (mathematical) model which relates micro-level interventions of interest to macro-level organizational productivity measures. Such a model would also include other important variables, operating at various organizational levels, which influence organizational productivity. Such a model might be thought of as a path, causal, or structural equation model (Loehlin, 1987). This kind of model may be visualized as a series of boxes, reflecting variables in the model. These boxes are connected by various arrows. In this visualization, each arrow might represent one or more equations linking pairs or sets of variables at different levels of abstraction. The major variables that influence outcome variables are represented in the equation set and are structured so that output variable (e.g., organizational productivity) estimates are appropriately sensitive to changes in the various input variables.

This sort of model will permit users to estimate impacts on outcome variables of interventions at lower levels of abstraction (e.g., at the individual level) while holding other variables constant. Such a model can also be used to study the relative sensitivity of outcome variables to the various types of input variables. This sort of sensitivity analysis can be used to determine which organizational interventions, if any, are likely to have practical value in light of the many other factors that influence outcome variables.

How might such a model be built? One approach is to estimate all model parameters simultaneously from a single set of data using a latent variable structural equation modeling procedure (e.g., Loehlin, 1987). However, this approach is often not feasible for real-world applications. Instead, it may be necessary to construct the integrated model from various submodels. Such submodels themselves may not be statistically estimated from integrated data or, if estimated in this way, may not be appropriate for the overall purpose of the model. This issue is discussed in more detail below.

The purpose of this paper is to describe the Training Decisions System (TDS) (Vaughan, et al., 1989; Mitchell, et al., 1992), an organizational simulation modeling system which meets the requirements described here. The TDS illustrates how a useful organization intervention analysis model can be developed under a variety of real-world constraints.

The Training Decisions System

The overall purpose of the TDS is to support strategic manpower, personnel and training (MPT) planning for specialties (occupations) within the Air Force. The TDS meets this objective by estimating important organizational impacts associated with structuring training for an occupation in various different ways, so that tradeoffs associated with different approaches to meeting task training requirements (e.g., mixes of classroom, lab, and on-the-job training) can be studies. The overall objective in building the TDS model was to take into account the key drivers of training resource requirements and costs, including non-training-related factors.

The first issue in building the TDS concerned organizational productivity. Conventional approaches to studying training impacts on organizations would relate training directly to organizational productivity. This approach has been attempted in conventional training utility analysis studies (Cascio, 1989; Mathieu and Leonard, 1987). However, relating training to overall organizational productivity can be difficult. In noncommercial settings, such as the Air Force, it can be difficult even to define the theoretical organizational productivity construct in a way that is applicable to entire occupations. In the TDS, we avoided this problem by fixing organizational productivity at a constant level, and then estimating training resource requirements, costs, and capacities required to achieve the fixed productivity for various training scenarios. In this way, influence of training on meaningful organizational variables (e.g., operating budgets and resource requirements) is estimated, permitting identification of the least expensive way of meeting training requirements that is consistent with training resource availability constraints.

Table 1 presents the key TDS model inputs and outputs. As may be seen, the key outputs include estimated training quantities, for both formal training and on-the-job training (OJT). Outputs also include estimated resource requirements, both labor and nonlabor, as well as costs and capacities, for both formal training and OJT. Inputs include a formal training structure scenario, which includes specifications of training hours on each task using each major training delivery method (e.g., classroom, self-study, laboratory) for each formal course. Inputs also include a job structure scenario. In general, training requirements are driven by job task performance requirements. Furthermore, the structure of jobs, the previous training and experience that people bring to jobs, the geographic distribution of jobs, and other job-related factors can be training cost drivers. Thus, the model

incorporates these job-related variables, as well as training-related variables.

INPUT	OUTPUT
Utilization and Training Pattern Model	
1. Jobs 1a. Task Content 2. Formal Training Courses 2a. Task Content 3. Transition Probabilities	1. Formal Training Quantities (by course) 1a. Numbers of Trainees 1b. Training Hours by Task 2. On-the-job Training Quantities (by job and task) 2a. Numbers of Trainees 3. Job Flow Quantities
Resource, Cost, and Capacity Model	
1. Training Quantities (from Job and Training Flow Model 2. Training Resource Requirement Model Parameters 3. Cost Factors 4. Student-Teacher Ratios 5. Training Resource Availability Quantities	1. Training Resource Requirements 1a. Formal Courses 1b. On-the-job Training 2. Training Costs 2a. Formal Courses 2b. On-the-job Training 3. Training Capacity Estimates 3a. Formal Training Sites 3b. On-the-job Training Sites

Table 1. Key TDS Model Inputs and Outputs

Figure 1 illustrates the TDS modeling process (Mitchell, et al., 1992:32-35). As may be seen from this figure and from Table 1, the TDS model includes two major submodels. The first submodel, the Utilization and Training (U&T) Pattern Model, estimates numbers of people taking each formal training course defined in a scenario, as well as the additional OJT required to meet all job task performance requirements estimated for a scenario. This U&T pattern submodel, in turn, has two major components. The first of these, the U&T Pattern Simulation, is a discrete event digital simulation of the flow of individual airmen through jobs and training courses over their careers. This model estimates organizational-level job and training flows, based on simulation of career paths followed by individual airmen. The second U&T pattern model component estimates training quantities required to meet the job flows from the simulation model. For formal training, this simply involves counting numbers of simulated airmen who take each course. For OJT, training quantity estimation is more complex. OJT estimates are made for each entry of an airman into a new job in the simulation. An OJT estimate is done for each group (module) of tasks assigned to the airman in the new job. These task module-specific estimates take into account previous training received by the airman on the tasks. The estimates also make use of task module-specific allocation curves, or learning curves, to estimate entering proficiency, and the OJT hours required for a particular simulated airman to reach full proficiency on a task module. The individual OJT estimates are accumulated to arrive at average OJT quantities required by the scenario.

The Resource Cost Submodel estimates training resource quantities, costs, and capacities required by the training quantities from the U&T Pattern Model. This submodel, in turn, has three components. The first is the resource requirement component. In this component, task and training delivery method-specific linear functions are required to estimate the quantity of each resource required by the training quantities. Resources include labor, both student and instructor, and equipment. Then factors are applied to estimate costs required by selected resources (e.g., labor costs). Capacity analyses then compare resource requirements to resource availabilities at various geographic locations. Thus, the individual task performance requirements and training events are translated into overall costs and capacities at the various training sites required for a job and training

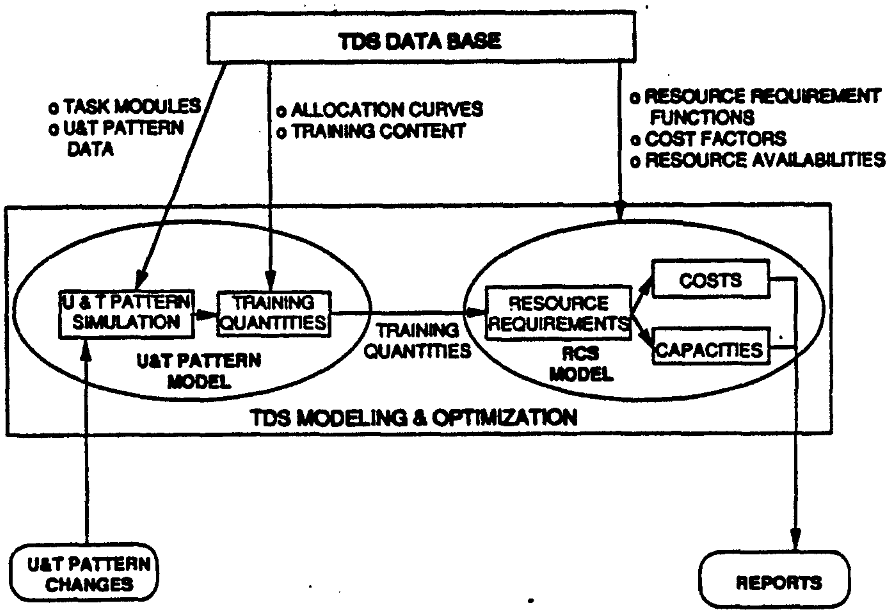


Figure 1. TDS Modeling Process

structure. This process takes into account previous training and experience as well as the various cost-related advantages and disadvantages of both formal training and OJT. For example, the model takes into account the economy of effort of formal training, in which many students can share instructors and equipment, as well as the diseconomies of such factors as travel required to bring students to the formal training location and of the fact that all students in a class are trained on all tasks covered in the class, even of only a subset of the students will be required to perform a particular task. Similarly, the model considers the economies of OJT, which requires no travel and trains students only on tasks they will be required to perform, as well as the diseconomies associated with nearly one-to-one student-instructor ratios and instructional inefficiencies in OJT.

A TDS model for an organization or occupation requires a great deal of data, about the job structure, training structure, and tasks to be modeled. These data come from many different sources. Some of these sources are existing data bases, including the occupational survey/CODAP job analysis data base, and various manpower, personnel and training (MPT) data bases. In addition, much of the data come from subject-matter experts' (SMEs') judgments. As the above discussion suggests, a TDS model for an occupation is built in pieces. Various types of model parameters are estimated separately, from separate data sources. These sets of parameters are then combined to form a complete TDS model. These model pieces are smaller even than the components described above. For example, in the U&T pattern simulation, jobs and task content, training courses and task content, and transition probabilities must be estimated separately (Mitchell, Yadrick, & Bennett, in press). Furthermore, various different transition probability subsets must be estimated separately. In many cases, multiple data sources must be used for estimating a particular parameter class. Furthermore, parameters may not be obtained using strict statistical estimation procedures. One reason for this concerns issues in combining inconsistent data from multiple sources. Another stronger reason concerns the generalizability and application of model results. Historical data often reflect the state of an occupation several years ago. However, the model results are used to describe the current situation and extrapolate to the future. Thus, it is often necessary to adjust historical model parameter estimates to reflect recent and expected changes.

The baseline model against which any alternative training and job configuration is to be evaluated is represented in data describing current job and training programs. Proposed changes are represented by modifying the descriptive data, and then running a new simulation. Results are compared in terms of changes in the total annual cost of training programs as well as in the capability of representative units to conduct such training (Ruck, 1989).

SMEs' judgments play a very important role in TDS model parameter estimation. In particular, the learning curves and the training resource requirement estimation functions are based on SMEs' judgments (Vaughan, Perrin, Ruck, & Bennett, 1991). This illustrates another important point about building organizational

intervention analysis models. The TDS model requires that costs be related to individual task training. This permits training on particular tasks to be modified in alternative scenarios. Conventional accounting principles would meet this requirement by allocating total training costs to individual tasks by simple proportioning rules-of-thumb. However, such rule-of-thumb cost allocation is not sensitive to key training cost drivers, such as those mentioned above. In general, traditional accounting approaches for allocating costs (e.g., overhead costs) are often not useful for decision support, because they are not sensitive to true cost drivers. Activity-based accounting approaches (Kaplan, 1988) have been developed for these reasons. In activity based accounting, SMEs' judgements and other nontraditional data sources are used to estimate relations between true cost drivers and resulting costs. These relationships may be less precise, but are more accurate, and better meet decision support needs. We have used this activity-based accounting approach in the TDS by using SMEs' judgments to estimate relationships between task training quantities and required resource quantities, to support training cost and capacity estimation.

Conclusion

A major difficulty in planning and evaluating organizational interventions concerns the weak empirical relationship between micro-level interventions and macro-level organizational outcome or productivity variables. This weak relationship reflects the many uncontrolled moderating and external variables which influence organizational productivity and are uncontrolled in evaluation studies. A practical solution to this problem involves building a mathematical model which relates macro-level outcome variables to their causes at various organizational levels. This sort of model would permit examination of micro-level organizational intervention impacts on macro-level outcome variables while holding constant other key factors.

The Training Decisions System (TDS) exemplifies such a multi-organizational-level model. The TDS relates micro-level training interventions to organizational-level training costs and capacities. In the TDS, organizational productivity is held constant, thus avoiding one of the most difficult aspects of building such a model. Approaches used in building the TDS are discussed, including separate construction and validation of various model parts and use of activity-based accounting approaches, and SME judgments.

References

Cascio, W. F. (1989). Using utility analysis to assess training outcomes. In *Training and Development in Organizations* (Goldstein, I.L., Ed) . San Francisco, CA: Jossey-Bass.

Kaplan, R. S. (1988, January-February). One cost system isn't enough. *Harvard Business Review, 66*: 61-66.

Loehlin, J. C. (1987). *Latent Variable Models.* Hilldale, NJ: Lawrence Erlbaum Associates.

Mathieu, J. E., Leonard, R. L. (1987). Applying utility concepts to a training program in supervisory skills: A time based approach. *Journal of the Academy of Management* 30:316-335.

Mitchell, J. L., Yadrick, R. M., & Bennett, W. R. (in press). Estimating training requirements from job and training models. *Military Psychology.*

Mitchell, J. L., Vaughan, D. S., Knight, J. R., Rueter, F. H., Fast, J., & Haynes, W. R. (1992, June). *Training decisions technology analysis* (AL-TP-1992-0026). Brooks AFB, TX: Technical Training Research Division, Armstrong Laboratory, Human Resources Directorate.

Ruck, H. W. (1989). Symposium. The Air Force training decisions system: Modeling management and policy. *Proceedings of the 31st Annual Conference of the Military Testing Association* (pp. 563-593). San Antonio, TX: Air Force Human Resources Laboratory and the USAF Occupational Measurement Center.

Vaughan, D. S., Mitchell, J. L., Yadrick, R. M., Perrin, B. M., Knight, J. R., Eschenbrenner, A. J., Rueter, F. H., & Feldsott, S. (1989, June). *Research and development of the training decisions system* (AFHRL-TR-88-50). Brooks AFB, TX: Training Systems Division, Air Force Human Resources Laboratory.

Vaughan, D. S., Perrin, B. M., Ruck, H. W., & Bennett, W. R. (1991, October). Estimating learning curves from job analysis data. *Proceedings of the 33rd Annual Conference of the Military Testing Association*. San Antonio, TX: Human Resources Directorate, Armstrong Laboratory and the USAF Occupational Measurement Squadron.

ODMARS - AN OCCUPATIONAL ANALYSIS PRODUCTIVITY ENHANCEMENT TOOL

Squadron Leader John S. Price
Royal Australian Air Force

Background - Occupational Analysis in the RAAF

1. Knowledge of the nature and requirements of jobs is a prerequisite for effective personnel management in any organization, whether large or small, public or private. The RAAF, as a large public organization with a crucial military mission, has a vital interest in determining precisely the work performed by its airmen, airwomen and officers in each of its many employment categories. This information about jobs is needed for personnel selection and recruitment, determination of trade specifications, formulation and validation of training course content, assessment of job satisfaction levels, identification of occupational health and safety problems, and for other specific personnel management purposes.

2. The RAAF has employed a task inventory method of occupational analysis for almost two decades. Task performance data is collected directly from job incumbents using surveys that seek to identify the tasks they perform in their respective jobs and the relative amount of time they spend performing each task. To assist with training decisions, senior members of an occupational group also provide training emphasis and learning difficulty ratings on all tasks performed by members of the particular target population being surveyed. The survey also collects from job incumbents biographical and background information relevant to the resolution of issues specific to the occupation, such as age, length of service, educational attainment and job satisfaction level. Basic processing of this survey data is accomplished on the Defence mainframe computer located in Canberra, using a suite of programs known as CODAP (an acronym for Comprehensive Occupational Data Analysis Programs). CODAP manipulates survey data and produces reports that aid in the analysis of the occupation. The RAAF Occupational Analysis Cell (PE1) within Headquarters Training Command (HQTC), located in Melbourne, develops the surveys, analyzes the data and produces occupational survey reports. Optical scanning and processing of survey responses using CODAP are performed by the Headquarters Australian Defence Force Occupational Analysis Cell in Canberra on behalf of the RAAF and other government agencies.

Occupational Data Manipulation and Reporting System (ODMARS)

3. The Occupational Data Manipulation and Reporting System (ODMARS) is an integrated set of application programs which accesses commercially-available relational database software. It has been conceptualized, designed, programmed and implemented to meet organizational needs for increased productivity, efficiency and quality in the occupational analysis process, and also to provide new capabilities in the use of occupational data by the RAAF.

4. ODMARS consists of three functional modules, each with various applications. It has electronic interfaces with CODAP data products via floppy disc and modem. ODMARS has been developed to run on either of two computer operating systems: UNIX and MS-DOS. The UNIX version is for use on the RAAF HQTC local area network, while the MS-DOS version runs on a portable notebook personal computer. Two versions of commercially- available relational database software are used: INFORMIX SQL for UNIX and INFORMIX 4GL for MS-DOS.

5. Before the development of ODMARS in 1991/92, all handling of CODAP data within PE1 was via hardcopy printouts. All manipulation of data in the analysis process either relied on ordering additional CODAP products from the Defence mainframe (with consequent delays in delivery) or was performed manually. In addition, all production of data displays for occupational survey reports was accomplished via keyboard entry (on typewriters until the early 1980s, and on wordprocessing systems since). ODMARS has automated most of the in-house data manipulation workload and eliminated much of the requirement for keyboard entry of task statements and accompanying data.

Task Inventory Development Module

6. The **Task Inventory Development** module of ODMARS is the tool by which the heart of any occupational survey (that is, the list of job tasks) is developed. Task statements are entered into the database as they are composed. The database structure has been designed to permit automated production of three distinct versions of the developing task inventory at any stage in its development. The three versions are Noun (entire task inventory listed alphabetically by noun), Verb (alphabetically by verb) and Inventory Order (listed alphabetically by verb, but within Duty Area groupings). These three versions are used concurrently by the PE1 inventory developer during small group facilitation sessions with selected members of the occupation under study. They serve to enhance the comprehensiveness and accuracy of the task inventory. Task statements need only be keyed in once, directly into the database. Amendment or deletion of tasks is accomplished by simple means, again directly on the ODMARS database. This module also generates the final camera-ready version of the task inventory for the survey booklet as well as the electronic version for uploading into CODAP. Inventory development sessions are mostly carried out in or near the workplace of job incumbents across the RAAF: that is, away from HQTC. The notebook PC version of this module enables the inventory developer to use these powerful applications in the field, thus negating the need to return to HQTC for successive updates of the developing inventory.

Data Reports Module

7. The **Data Reports** module can be thought of as a simpler, user-friendly, versatile, stand-alone version of some elements of the CODAP software. For example, this module accepts raw CODAP task data products (such as PRTFAC) in electronic format and generates displays of significant or representative tasks (with accompanying selected data) for specific target sub-populations such as rank groups or job types. These reports assist the analysis process and also serve as camera-ready data displays for inclusion in the occupational survey

report (again, with consequent elimination of much of the typing workload). Reports from this module can draw on various data elements to achieve a range of new applications and satisfy special or one-of requirements. For example, occupational data can be tailored to facilitate reviews of RAAF Trade (Competency) Standards, or for cross-study comparisons where amalgamation or job redesign is an issue.

Decision Support Module

8. The Decision Support module applies more complex programming logic to occupational data and accepts direct input of policy or expert judgements to generate detailed products that support RAAF decision-making. The one currently complete application within this module serves to support training content decisions. Occupational data are processed in a way that reflects RAAF training policy to produce, in an automated fashion, a training priority indicator for every task in an occupation. This helps the RAAF to determine new training requirements or validate existing training against job needs. The power of this module lies in its ability to automatically and quickly process vast amounts of job performance data and expert ratings, and to highlight the need for, and also to capture, additional judgements on tasks during the decision-making process. In addition, embedded algorithms are easily modified to reflect changing training philosophies or policies, or to enable investigation of "what if" scenarios. This module performs in minutes a function that formerly took weeks in manual mode, and now has increased levels of sophistication, versatility and capability.

Project Implementation and Impact

9. Occupational Analysis (OA) is a fundamental process in the RAAF personnel and training system. The RAAF is a leader in this field in Australia, mainly by virtue of the Officer Exchange Programme with the United States Air Force. For the last two decades, RAAF officers have worked at the USAF Human Resources Laboratory in San Antonio, or its predecessor or successor organizations. USAF officers have formed an integral part of the RAAF Occupational Analysis Cell in Melbourne. Traditional drawbacks of the task inventory method of occupational analysis have been its manpower-intensive nature and the long project time required to complete a study. Until recently, RAAF OA project durations of the order of twelve months have been normal. Any reduction in this timescale greatly enhances the usefulness of occupational analysis in providing data to support personnel initiatives in 'real time'.

10. HQTC-PE1 conceptualized ODMARS and developed detailed system performance parameters. The programming, testing, implementation and documentation of the system were accomplished using in-house resources. ODMARS technology represents a quantum leap in efficiency and improved customer service. By automating the data flow across the whole OA process, ODMARS has reduced typical project timescales to less than six months, improved the accuracy, comprehensiveness and quality of reports, and provided valuable new data products. These improvements have been accompanied by actual reductions in clerical support effort.

11. All RAAF occupations, technical and non-technical, airman and officer, have either recently been, or are currently undergoing, review for structural efficiency. Also, numerous entire RAAF work centres are being assessed to determine whether contractor support is a more efficient option than retaining uniformed manpower. RAAF OA data have been collected, analyzed and used to support decisions in many of these areas. By facilitating an increase in responsiveness to urgent organizational needs, ODMARS has enabled the RAAF OA cell to be relevant and effective in a time of massive and fundamental structural change.

12. ODMARS continues to evolve as new applications are developed, and further considerable productivity, efficiency and effectiveness payoffs are likely. The system has recently been 'exported' to the Royal Australian Navy Occupational Analysis Cell and the Headquarters Australian Defence Force Cell, both in Canberra. In August 1992, HQTC-PE1 was awarded a Letter of Commendation from the Chief of the Australian Defence Force and Secretary of the Department of Defence in recognition of the creation of ODMARS.

Differences in Content Validity Ratings Across Air Force Career Fields and Grade Levels

Trina K. Mayhill
United States Air Force Academy

Johnnie C. Harris
Daniel W. Schuette
Paul P. Stanley II
USAF Occupational Measurement Squadron

BACKGROUND

Enlisted members of the US Air Force compete for promotion to the grades of staff sergeant through master sergeant (E-5, E-6, and E-7) under the Weighted Airman Promotion System (WAPS). One factor used in WAPS is the member's score on a Specialty Knowledge Test (SKT). SKTs are 100-question multiple-choice tests which measure enlisted members' knowledge of their particular Air Force specialty. The tests are written at the USAF Occupational Measurement Squadron (USAFOMS) by teams of senior noncommissioned officers (NCOs) on temporary duty to develop tests for their respective specialties, under the direction of Squadron test development psychologists.

In May 1990, USAFOMS integrated the use of content validity ratings (CVRs) into the test development process. CVR forms are filled out by the SMEs, who then use the results in the question-writing process to help evaluate each test question in terms of its relationship to success in the specialty. The forms are based on the work of C.H. Lawshe (1975), who described content validity as the degree of overlap that exists between the sampled knowledge domain on a particular test and the knowledge necessary for performance of a particular job. This is consistent with Cronbach's (1971) statement that "Content validity is evaluated by showing how well the content of the test samples the class of situations about which conclusions are to be drawn."

Lawshe's method comprised a panel of SMEs who independently rated each question on a test according to the following scale:

"Is the skill (or knowledge) measured by this question:

___Essential (2)
___Useful but not essential (1), or
___Not Necessary (0)

for successful performance on the job?"

The question was modified for use in USAFOMS so that the "specialty," rather than the job, was the frame of reference for the raters. This was necessary because each test being rated was applicable to an entire Air Force specialty, which generally encompasses a range of related jobs. Two different forms were used for gathering validity ratings, as each specialty generally has two different SKTs, one test for promotion to E-5, and a separate test for promotion to E-6/7. The knowledge required for performance in one grade of a specialty may be considerably different from that required for performance at a different grade. The two forms allow the SME to rate

each question on the SKT on a scale which quantifies its validity of content and which applies specifically to the grade in question.

Since a rating of "2" on the CVR form means the rater considered the knowledge covered by the question to be essential to performance in that specialty, the USAFOMS point of view was that questions identified as such possessed adequate content validity for inclusion on the SKT. A rating of "1 " was, likewise, considered a nonproblem, since this means that the question was considered to be useful, if not essential. Researchers focused their attention on those questions rated "0," meaning that the SME considered that question unnecessary for performance in a specialty. Such questions were therefore considered to be suspect in regard to content validity.

Since there has been little research conducted in respect to CVRs, questions arise as to what assessments can be made based on the content validity of tests as determined by CVRs and how the CVRs should be used during test construction. This paper will explore both areas. In particular, are there significant differences in the average number of test questions that are given "zero" CVRs across grade level(s) and groups of similar specialties? Differences in the content validity of SKTs or groups of SKTs could have a negative impact on the accuracy of WAPS and, therefore, the proper evaluation of enlisted personnel for promotion opportunities. When CVRs were first implemented in USAFOMS, USAFOMS psychologists were told that specific guidelines on how to use the forms would not be provided until after a trial period during which they could try various ways to use the forms. It is now an appropriate time to make this assessment. It was felt that interviewing USAFOMS psychologists would provide useful input for establishment of guidelines for improved CVR form use.

PURPOSE

There were two main purposes for this research. The first was to determine if there were significant differences in the number of zero CVRs across grade level(s) and groups of similar specialties. The second purpose was to determine project psychologists' attitudes toward CVRs and the use of CVR forms during test development.

Determination of the differences in the average number of test questions receiving zero CVRs across grade level(s) and groups of specialties would give more information on the implementation of CVR forms in USAFOMS. Information on the consistency of content validity across various SKTs would help USAFOMS to improve the homogeneity of the various tests, therefore increasing the adequacy of the testing of all enlisted personnel regardless of grade level or specialty. Knowledge of USAFOMS psychologists' attitudes and present use of CVR forms could also allow more appropriate use of the CVR forms during test construction and aid in the establishment of guidelines for future CVR form use.

RESEARCH SAMPLE

The sample comprises SME content validity ratings from 48 of approximately 200 tests which had been rated. In order to ensure that the ratings examined were comparable across grades and specialties, this sample was selected as follows:

1. All specialties were grouped by career field (The first two digits of a five-digit Air Force specialty code identifies the career field, a grouping of specialties which are similar in terms of the skills and knowledges required for performance).

2. Specialties with fewer than four complete SME ratings were eliminated from consideration.

3. Specialties which did not have both an E-5 and an E-6/7 SKT were eliminated from consideration.

4. Career fields in which fewer than three specialties met the above criteria were eliminated from consideration.

The career fields remaining as a result of this process of elimination were as follows:

11xxx - Aircrew Operations
27xxx - Command Control Systems Operations
30xxx - Communications-Electronics Systems
45xxx - Manned Aerospace Maintenance
55xxx - Structural/Pavements

INTERVIEW SAMPLE

All fully qualified test psychologists (TPs) and test management psychologists (TMPs) within USAFOMS were interviewed regarding their use of and attitudes toward CVR forms throughout the test development process. There were a total of 23 USAFOMS psychologists interviewed, 8 TMPs and 15 TPs.

ANALYSIS

It was felt that counting questions rated as zero by only one SME might give an overestimate of test content validity problems. When only one SME of the group rates a question as unnecessary, there is some chance that this judgment may simply result from a relatively narrow experience base. When two or more SMEs give an item a zero rating it is more likely that the item does indeed lack content validity in a particular specialty. For this reason, the data were analyzed twice, once counting questions rated as zero by only one SME and again counting those questions rated as zero by two or more SMEs.

A two-way ANOVA with alpha set at .05 was used to test for the possible differences in the average number of questions receiving zero CVRs in respect to the grade level(s) and groups of specialties associated with each SKT. A two-way ANOVA was performed for the two categories of: 1) only one SME rating a question as zero and 2) two or more SMEs rating a question as zero. There were three null hypotheses to be tested for each category. These were: (H01) there are no differences in the number of zero ratings between the SKT grade levels of E-5 and E-6/7, (H02) there are no differences in the number of zero ratings between the various groups of specialties sampled, and (H03) there are no interactive effects after the main effects of grade level and career field have been removed.

RESULTS

Two-way ANOVAs were performed for each of the categories. The results are presented in Tables 1 and 2. In the first two-way ANOVA, we failed to reject all three null hypotheses. There were no significant differences in the zero ratings across the grade levels of the SKTs ($F_{1,38} = 1.36$, $p < .05$) or the groups of specialties ($F_{4,38} = .95$, $p < .05$). There were also no interactive effects between the grade levels and groups of specialties ($F_{4,38} = .12$, $p < .05$). This ANOVA summary is presented in Table 1.

Table 1.
Summary Table for Two-Way ANOVA
With Disproportionate Cell Frequencies
(Zero Ratings by Only One SME)

Source	SS	df	MS	F	Fcv
Rows	51.12	1	51.12	1.36	4.10
Columns	143.07	4	35.77	.95	2.63
Interaction	18.11	4	4.53	.12	2.63
Within		38	37.64		

In the second two-way ANOVA, testing the category with two or more SMEs rating a question as zero, we also failed to reject all three null hypotheses. There were no significant differences in the zero ratings across the grade level(s) of the SKTs ($F_{1,38} = .04$, $p < .05$) or the groups of specialties ($F_{4,38} = 1.09$, $p < .05$). There were also no interactive effects between the grade level(s) and groups of specialties ($F_{4,38} = .04$, $p < .05$). The summary of this two-way ANOVA is presented in Table 2.

Table 2.
Summary Table for Two-Way ANOVA
With Disproportionate Cell Frequencies
(Zero Ratings by Two or More SMEs)

Source	SS	df	MS	F	Fcv
Rows	.81	1	.81	.04	4.10
Columns	89.10	4	22.28	1.09	2.63
Interaction	3.66	4	.92	.04	2.63
Within		38	20.50		

Following the statistical analysis of the content validity ratings, USAFOMS psychologists were interviewed. They were asked eight questions seeking the attitudes and knowledge of each psychologist in respect to content validity and their use of the CVR forms during the test development process. The results of the interviews were categorized by common responses and the frequency of each response was recorded.

DISCUSSION

Failure to reject the null hypotheses means that there appear to be no differences in the way test questions are related which can be attributable to either the career field in which the test exists or the grade level at which the test is aimed. From a practical standpoint, this lends indirect support to the way in which USAFOMS uses the ratings (*i.e.*, ratings are used in the same manner and given the same weight in making test development decisions regardless of the grade level or career field involved). In other words, management practice regarding the CVRs is, at least, not inconsistent with the findings.

Follow-on research will be necessary to determine whether this relationship exists throughout the rest of the tests. In addition, future research will be aimed at interrater reliability, because one of the problems from the statistical perspective is that the ratings are rendered by such a small group of experts.

The interviews conducted with the TMPs and TPs of USAFOMS provided interesting insights into their attitudes and use of the CVRs throughout the test development process. The psychologists expressed the desire to know the current and future status of the CVR forms. They were told upon the implementation of the forms that there would be a trial use period after which guidelines would be established on the use of the CVR forms. It appears the psychologists now feel that if the CVR forms are to remain in use, guidelines should be developed. Guidelines on the appropriate use of CVR forms will now be established to promote standardization in test construction.

Of the interview sample, 48% held positive or neutral attitudes toward the use of CVR forms, while 52% held negative attitudes. With some exceptions, TPs generally held positive attitudes toward the use of the CVR forms during test development. TMPs, on the other hand, generally held negative attitudes toward use of the forms. One reason stated for this view is the number of SMEs involved in the test development process. TMPs are associated with minor revision projects, which usually consist of only two SMEs. Their opinion was that the CVR forms are redundant for use with a small number of SMEs because the content validity evaluation and discussion occurs spontaneously, without the CVR forms. The TPs, however, are associated with major revision projects, which involve four to seven SMEs. Most TPs expressed that CVR forms are very helpful to prompt SME question discussion with a sizable group where discussion may be difficult. Consideration will be given to making CVRs mandatory for major revisions and optional for minor revisions.

Questions were directed to the USAFOMS psychologists concerning their knowledge of content validity and its use as a strategy for test validation. Most were well aware of the concept of content validity, but the question referencing the strategy of content validity for test validation created some confusion. Training in this area will be provided to make the issue of content validity as a strategy and its relationship to the USAFOMS mission more clear. In addition, the new guidelines will mandate a general explanation of content validity to the SMEs upon beginning the CVR form use.

It was also mentioned during the interviews that a question rating scale with a wider range may be more useful. A scale from 0-4, instead of 0-2

might make it easier for the SMEs to decide more exactly the amount of content validity they feel a question possesses. On the other hand, this change may confuse an already difficult to quantify concept. Further research into the use of a five-point scale, as opposed to the three point scale in use, may be needed.

Further research is recommended to investigate the number of questions receiving zero ratings which continue to be used in testing. The question also arises as to whether questions receiving low content validity ratings continue to be used. Investigating these topics may provide more insight into the usefulness of CVR forms.

REFERENCES

Cronbach, L. J. Test Validation. ***Educational Measurement***, 1971, 443-507.

Lawshe, C.H. A Quantitative Approach to Content Validity. ***Personnel Psychology***, 1975, 28, 563-575.

Perry, C.M., Williams, J.E., and Stanley, P.P. Implementation of Content Validity Ratings in Air Force Promotion Test Construction. ***Proceedings of the 32nd Annual Conference of the Military Testing Association***, 1990, 235-240.

USAF Occupational Measurement Squadron ***Handbook for Construction of the SKT and Associated Tests***, 1988.

RATIONALE FOR COGNITIVE TASK ANALYSIS OF TACTICAL DECISION MAKING

JOHN A. MODRICK
Systems and Research Center, Honeywell Inc.

A method for cognitve task analysis (CTA) that is well-defined, standardized, validated and accepted does not yet exist. Approaches to cognitive analysis and analyses of tasks has been reported, providing concepts and CTA methods which can be validated over a series of applications. However, they have one or more drawbacks for analyzing tactical decision making (TDM): limited to deriving requirements for training and training equipment, deal with maintenance rather than operating tasks and jobs, or oriented toward extracting knowledge from experts for the design of expert systems. They were not designed to meet the broad range of information requirements needed to address the spectrum of issues in manpower, personnel, and training (MPT).

The purpose of task analysis is to provide data and information about job tasks to support decisions in equipment design and MPT. Uses of task analysis data include interface design and evaluation, performance evaluation, personnel and manning requirements, training support requirements, decision aiding, and preparation of technical manuals.

The proposed rationale for CTA is based on an integration of ideas and perspectives from recent research on naturalistic decision making and the nature of expert performance and expertise. They provide a conceptual framework for decision making in dynamic situations with uncertainity, ambiguity, and partial information.

Cognitive And Conventional Task Analysis

We use the following charcteristics as the criteria for cognitive activities: 1) Use of, or dependence on, knowledge organized into integrated, conceptual structures to support performance, 2) Use of mental models, 3) Some amount of covert information processing inferred from overt behaviors, and 4) Existence of a goal structure in terms of which the execution of skilled behavior is generated, controlled, and adapted to situational conditions that affect goal-attainment. A cognitive task analysis must provide information descriptive of the covert decision processes, knowledge structures, mental models, and goal structures.

Job behaviors are described in conventional task analysis in terms of the overt behaviors required of a proficient operator/user in performing the tasks in an operational environment. The tasks are treated as well-structured procedures organized into cue-response sequences of simple actions. Supporting and enabling skill and knowledge are inferred from these actions. However, the elements of knowledge are treated as a set of independent items. Although task analysis is part of a larger domain of job analysis, its principal use in human factors has been to determine the requirements for training and training equipment to support a system (Gagne, 1974). The methods for deriving training requirements have been augmented with the procedures for Instructional System Development (ISD).

Task analysis should be driven by the data needed during system design. The information required is determined by 1) the design goals for which the data will be used and 2) the cognitive processes underlying operational behaviors. The processes must be inferred from overt behaviors interpreted in the context of a decision process model. A challange in developing CTA is to determine what information can be obtained by retrieval from available sources and how to generate the remainder by analysis, inference, or invention.

Tactical Decision Making

TDM designates the activities of planning and executing a course of action in a dynamic, ill-structured, ambiguous situation. Cognitive activities, such as TDM, are goal-directed and procedural in execution. TDM entails complex activities of situation assessment, action selection, and monitoring. Situation assessment requires an active, expectancy-driven seeking of information to deveop working hypotheses about the state of the world (Kirschenbaum, 1992). These activities in turn require extensive covert processing, greater cognitive than psychomotor demands, and dependence on knowledge structures and mental models.

Several compatible views of TDM can be integrated into a more complex conceptualization. In naturalistic or recognition-primed decision making (Klein and Klinger, 1991; Klein, 1990), for example, the decision maker chooses and initiates a satisficing action on the basis of situation recognition. He adapts the procedure to the conditions of the situation and makes procedural changes as conditions change or he uncovers additional information.

The PARI model (Precursor, Action, Result, and Interpretation) treats decision making as adaptive problem solving in a dynamic, unstable task environment (Hall et al., 1990). Procedures are compiled in real time and adapted to the specific problem (Gott and Pokorny, 1987). These procedures must be made up, composed or improvised in real time in each situation. Procedures are not repeated since each problem situation in this dynamic, unstable task environment is novel. Therefore, *a priori* fixed procedure cannot be fashioned to apply to the situations; they must be formulated as behavioral sequences adapted to the specific problem at hand. We have tagged them with the label of *ad hoc* procedures.

The knowledge from which procedures are compiled is clustered into three knowledge categories: Procedural: Component behavioral routines that make up a procedure, System: Situational components to which procedures must be adapted, and Strategic: Processes and factors in deployment and control of procedural and system knowledge in execution of procedures. The content of strategic knowledge is goals, plans, and decision rules.

We extend the PARI architecture to TDM (Figure 1), adding Situational Knowledge to the constraints to which procedures must be adapted and changing Strategic Knowledge to Strategic Procedure Management. Strategic Procedure Management interacts with the tactical problem to manage the functions of selection, adaptation, monitoring and executive control of the improvised procedure in execution.

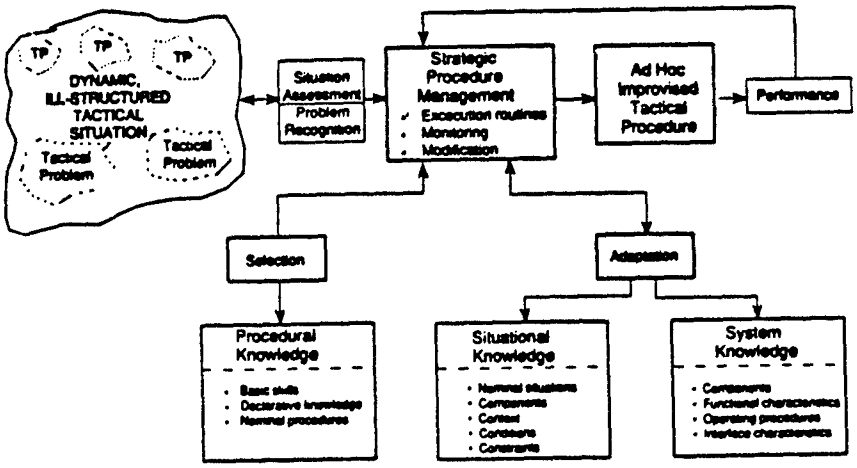


Figure 1. High Level Organization of Ad Hoc Cognitive Procedures

A tactical decision is also formulated as a Standard Operating Procedure that is an accepted, standard response to a recognized standard situation. It is a sequence of tactical actions that are instrumental to achieving an operational goal or subgoal. We represent the sequence of actions as a tactical procedure script (TPS) extending idea from Schank and Abelson (1977) and Abelson (1981).

TPS is formulated as goal-directed, semi-structured, cognitive procedures in emergent situations (Modrick et al., 1989; Parrish et al., 1988). A tactical action sequence is adapted to deviations from nominal tactical conditions, terrain, environmental features, and adverserial actions. A semi-structured procedure (Keene and Morton, 1978) is an action sequence with branching contingent upon outcomes, represented as a hierarchical structure of nodes. Emergent situations (Bougaslav and Porter, 1962) unfold during task performance, creating ambiguity and uncertainty. They require structuring incoming information, adapting situation perception, and changeing actions appropriately.

A challange for developing methods for CTA is to translate these categories into tactical things. Procedural knowledge is a chunking of actions with declarative knowledge to form cognitive skills. We need to find these chunks. Similarly, selection, adaptation, and strategic management are operators and rules that we need to uncover.

Framework For Cognitive Task Analysis

The CTA method consists of six phases: 1) Develop Inventory Of Tasks And Procedures In The Tactical Domain 2) Generate Behavioral Descriptions Of Tasks 3) Generate Task Data 4) Estimate Values For Task Attributes 5) Compile Task Data Base 6) Develop Procedures For Using Task Data To Address Design Issues In MPT.

Phase I. Develop Inventory Of Tasks And Procedures

The tactical domain is partitioned into tactical situations. The classification should be exhaustive of the operational activities in the domain. Each situation is partitioned into several problems. Within each problem one situation is chosen as the standard or nominal instance of the problem; there is also a set of variants on the nominal problem.

A standard tactical procedure, a TPS, is associated with the standard problem. It is the accepted way of dealing with the problem. A set of variant procedures is associated with a standard procedure. Each TPS is also represented as a tree structure of the sequence of actions, outcomes, and conditions; goal hierarchy; and state transition diagram.

Phase 2. Generate Behavioral Descriptions of Each Task

Verbal descriptions are written for the procedures and content on each task in the domain. Task descriptions include functions performed, behaviors of the person executing the tasks, and operational, tactical conditions that influence or constrain permissible actions. The descriptions must provide the information from which the task analysis can capture task parameters relevant to subsequent decisions on crewstation design, determine the knowledge and skill required, and identify contextual conditions of performance.

Phase 3. Generate Task Data

The data to be generated includes task characteristics from both conventional and cognitive analysis. Conventional task data includes frequency of performance, difficulty, performance times, criticality, time sharing, workload demand, competence level required, and knowledge/skill needed to understand/execute each task.

Glaser et al. (1991) have discussed the kinds of data and information needed from a CTA: Skills and abilities, selection rules, declarative knowledge used in adapting procedures, integrated structures coupling conceptual and procedural knowledge, networks of supporting knowledge, and mental models of tasks, equipment, situation, and actions.

Error types and consequences should be determined. Adams et al. (1991) have classified errors in cognitive management of multi-task systems as data misentry, misuse or misunderstanding of machine modes, misinterpretation of machine data/states, failures to take advantage of system assets and check system performance, inappropriate decisions to override/lock-up the system, and overloading/misappropriation of attention.

Reason (1990) has developed an extensive conceptualization of errors. He has classified errors as skill-based slips, rule-based mistakes, and knowledge-based mistakes.

Phase 4. Estimate Values for Task Properties

We have begun to cull articles that define attributes of tasks and procedures which maybe relevant to TDM. The complexity of procedural and perceptual-motor skills is characterized

by the attributes of memory demand, number of steps, sequence restrictions, feedback, time required, mental requirement, number and difficulty of facts required, and motor control required (Rose et al. 1985). They have been validated as predictors of forgetting.

Campbell (1988) defines four characteristics of cognitive complexity: multiple paths to a desired end state, multiple end states, conflicting interdependence among paths to multiple end states, and uncertain or probablistic links among paths and end states. Combinations of characteristics yield four task types: Decision, Judgment, Problem, and Fuzzy.

Tasks and procedures can be classified by type. Bloomfield et al. (1989) have developed a classification scheme for characterizing flight decisions, organized into five groups: situation attributes, decision functions, inputs, option characteristics, and resource demands. Decision functions include problem recognition, data acquisition, information evaluation, situation recognition, developing/evaluating options, and monitoring/adjusting options. Other activities can be added as justified, including hypothesis generation and testing, situation assessment, implementation of action, and recognizing key events.

Hammond et al. (1987) have differentiated between intuitive and analytical tasks on the basis of task characteristics. The task characteristics are defined by the nature, structure, and relationships among cues used in the tasks.

Phase 5. Compile Task Data Base

A Task Data File must be structured to meet the needs of the system designers as well as users and operators who must develop doctrine for deployment and procedures for operating the system. Organization of the task data should support retrieval for use in system design. Some means of correlation between data items, such as Hypercard, will be considered.

Phase 6. Develop Procedures For Using Task Data To Address Design Issues In MPT

Analytical routines and algorithms must be formulated so that the task data can be processed to yield the products needed in systems design. There should be analytical techniques for interface design, personnel requirements, training, etc. Some tools of this type exist as in training analysis and ISD. These methods are ancillary to or build on the CTA data rather than part of it.

Problems

The problems are 1) This approach is too time consuming and labor intensive to be affordable for all tasks. We need rules to identify tasks or decisions that require this extensive analysis. The problem is reminiscent of the conventional distinctions between gross and critical task analysis. 2) The technology cupboard is relatively bare. We do not yet have adequate tools. 3) It is difficult to get access to sufficiently rich scenario materials to use in working through the analysis. Notional war games and Tom Clancy's novels are not adequate.

Selected References

Adams, M. J., Tenny, Y. J., and Pew, R. W. (1992) Strategic Workload and the Cognitive Management of Advanced Multi-Task Systems. CSERIAC.

Bloomfield et al. (1989) Flight Decision Characterization: Draft Final Report F33615-89U/P6754, Air Force Systems Command, HSD, Wright-Patterson AFB, OH.

Campbell, D. J. (1988) Task Complexity: A Review and Analysis. Academy of Management Review, 13, 40-52.

Glaser, R., Lesgold, A., and Gott, S. (1991) Implications of Cognitive Psychology For Measuring Job Performance. In A. K. Wigdor and B. F. Green, Jr. (Eds) Performance Assessment For The Worikplace, 1-26, National Academy Press, Washington, D. C.

Gott and Pokorny (1987). The Training of Experts for High-Tech Work Environments. Ninth Interservice/Industry Training Systems Conference, 184-190. Washington, DC.

Hall, Gott, and Pokorny (1990) A Procedural Guide to Cognitive Task Analysis: The PARI Methodology. AFHRL Technial Paper, Manpower and Personnel Division, Air Force Human Resources Laboratory, Brooks Air Force Base, TX. In press.

Hammond et al. (1987) Task Characteristics That Differentiate Between Intuitive and Analytical Tasks. IEEE: Systems, Man, and Cybernetics, SMC-17, 753-770.

Kirschenbaum, S. S. (1992) Influence of Experience on Information-Gathering Strategies. Journal of Applied Psychology, 77, 343-352.

Klein, G. and Klinger, D. (1991) Naturalistic Decision Making. CSERIAC Gateway, II, 1-4, Wright-Patterson AFB, Ohio.

Modrick, Plocher, and Parrish (1989) Methodolgy and System Design for Performance Evaluation in Tactical Trainers. Proceedings, Sixth Annual Workshop on C2 Decision Aiding, NOSC, San Diego, CA.

Reason, J. (1990) Human Error. Cambridge University Press, Cambridge, MA.

Rose, Radtke, Shettel, and Hagman (1985) User's Manual for Predicting Military Task Retention. ARI Research Product 85-26, U. S. Army Research Institute for the Behavioral and Social Sciences, 5001 Eisenhower Avenue, Alexandria, VA

10/21/92

Training Emphasis Based on Post-Operation Desert Shield/Storm Occupational Surveys

Lawrence A. Goldman, Ph.D.
U.S. Total Army Personnel Command
Alexandria, VA 22332-1333

<u>Background</u>. Since the early 1980's, the Army Occupational Survey Program (AOSP) has collected, on a routine basis, Training Emphasis (TE) data from senior supervisors/managers for nearly all surveys of officer branches/ functional areas, warrant officer Military Occupational Specialties (MOS), and enlisted MOS. The major purpose of collecting TE data, supplementing information obtained from job incumbents, is to assist U.S. Army training schools in deciding which tasks should be trained - either by the proponent school with the responsibility of providing structured, formal training or by supervised on-the-job training. Army training essentially focuses on what the soldier will be asked to perform in a wartime environment. In essence, TE raters are asked to help determine critical tasks to be trained based on their knowledge and experiences as senior officers or non-commissioned officers (NCOs) in their particular field. However, before early 1991, there was no opportunity for the AOSP to collect and analyze task information based on actual wartime experiences in revamping personnel and training programs. Just after the conclusion of the air/ground war for Operation Desert Shield/Storm (ODS/S) in March 1991, the AOSP initiated an action with the objective of obtaining and quantifying the experiences of Army personnel in 75 "key" commissioned officer branches/functional areas and warrant/enlisted MOS who participated in ODS/S.

Inasmuch as there has never been any previous AOSP effort to collect and analyze the work of Army personnel performing warfighting jobs, this study sought to provide insight into the following areas:

(1) the correlation between TE ratings based on soldiers with actual experience in ODS/S versus those individuals who didn't participate in ODS/S.

(2) the kinds of tasks which best discriminated raters who participated in ODS/S versus those who didn't participate based on their TE ratings.

(3) the relative rankings of the TE ratings of those tasks isolated as the best discriminators for those soldiers who participated in ODS/S versus those who didn't participate.

<u>Methodology</u>. The results reported in this study were based on Army-wide sample surveys of senior officers in Branch 12 - Armor and in MOS 93C - Air Traffic Control (ATC) Operator. For Branch 12, four sets of 100 questionnaires each were distributed, in October 1991, to Armor officers who were asked to rate Armor Officer Basic Course (AOBC) and Armor Officer Advanced Course (AOAC) training. The sample sizes of Armor captains rating AOBC training were 44 and 34 for those who participated in ODS/S versus those who didn't. The samples of Armor majors rating AOAC training were 60 and 37, respectively, for those raters participating in ODS/S as opposed to those who didn't. With respect to MOS 93C, two sets of 80 questionnaires each were also administered in October 1991 to senior NCOs who were asked to rate the training for skill level one (entry level) soldiers. The sample sizes for ATC Operator raters participating in ODS/S versus those who didn't were 43 and 53, respectively.

The TE scale used for MOS 93C (as well as other studies of enlisted soldiers) had eight values; the value of "1" signified "cannot evaluate"; "2" indicated that the rater believed that the task was "not a SL1 task"; "3" indicated that no formal training was required for SL1; "4" indicated low TE thru "8" indicating a high TE value. A similar scale was used to collect TE ratings for Armor officers, with the major exception that there was no scalar value representing that a task was not an AOBC or AOAC task. Thus, average TE values were based on the values of "2" thru "7", with "2" representing low TE in AOBC/AOAC, and "7" representing high TE in AOBC/AOAC.

The Comprehensive Occupational Data Analysis Programs (CODAP) were used to process all the files for Branch 12 and MOS 93C. With respect to the TE files, the internal consistency of the data was maximized, reflected by two types of reliability, coefficients which were computed for each TE data file obtained from senior raters: (1) the average inter-rater reliability of a single rater and (2) the stepped-up reliability coefficient reflecting the overall group of raters for a particular TE file. The reduced TE files, raters with divergent rating policies having been removed, were then moved to become external data sets comprising the data files used as input for execution of the Statistical Package for the Social Services (SPSS). Initially, to examine the inter-correlations between the raters who were involved in ODS/S versus those who weren't, a Pearson correlation coefficient matrix was generated. Then, to determine those tasks which best discriminated between raters who participated in ODS/S versus those who didn't, step-wise discriminant function analysis was utilized. Tasks were entered as discriminators which maximized reduction of the Wilks' Lamda. To determine if the best discriminators of the TE ratings tended to be higher for those raters who were involved in ODS/S as opposed to those who weren't, the rank orders of the TE ratings for each of the best discriminators were obtained and difference scores obtained between those raters involved in ODS/S as compared to those who weren't.

Findings.

a. TE Reliability. In general, the R_{11} and R_{KK} values for the Armor AOBC raters and the AOAC raters were consistently very high. Moreover, these reliability values were essentially identical for those raters who participated in ODS/S and those who didn't while they were very close between the two sub-groups of Branch 12 majors rating the AOAC. Specifically, the R_{11} values for the captains rating the AOBC were each .52, the R_{KK} for the officers participating in ODS/S being .98 while that for the officers not participating in ODS/S being .97. Similarly, the R_{11} value for the majors rating the AOAC participating in ODS/S was .41 while that for the complementary group was .36; the corresponding R_{KK} values for these complementary groups of majors were .98 and .95, respectively. Pertaining to MOS 93C, the R_{11} values for the complementary groups were .26 and .28 while the R_{KK} values were .95 and .93, respectively, for the senior NCOs who participated in ODS/S versus those who didn't.

b. Inter-correlations. The Pearson correlation coefficient between the raters who were directly involved in ODS/S and those who weren't were all statistically significant and extremely high for Branch 12 and MOS 93C raters. For Branch 12 captains rating the AOBC, the coefficient was nearly unity (.99) between those who participated in ODS/S versus those who didn't, while the coefficient for majors rating the AOAC was also nearly unity (.98). For MOS 93C, the coefficient was .88 between those senior NCOs who participated in ODS/S and those who didn't participate.

c. Prediction of tasks best discriminating ODS/S experiences versus non-ODS/S participation. The goal of the use of stepwise discriminant function analysis was to determine the set of tasks which best predicted "correct" group membership based on ODS/S experiences as opposed to non-participation in ODS/S. That is, it was desired to isolate those tasks (predictors) which maximized the percentage of all cases (i.e., raters who either participated in ODS/S or who didn't) classified correctly.Table 1 displays those tasks which best achieved this objective for the Branch 12 OBC raters. Table 2 displays the best discriminating tasks for the Branch 12 OAC raters while Table 3 shows the best tasks for the MOS 93C raters. For each task selected in these three studies, the change in Wilks' Lamda is shown for each task variable entered. Each of these tables also displays the rank order of the average TE ratings for both complementary groups of raters (based on the total task inventory of 696 tasks for Branch 12 and 480 tasks for MOS 93C) and the difference between the rank orders for each task (Note: a positive difference indicated that the rank order of the TE ratings for those raters who participated in ODS/S was

higher).

As shown in these three tables, those tasks which best discriminated raters who participated in ODS/S versus those who didn't were as likely to reflect higher TE ratings by ODS/S participants as by those who didn't. Moreover, these complementary rankings were as likely to be relatively high as relatively low (i.e., in relation to the total number of tasks in the inventory). With respect to the Branch 12 OBC study (Table 1), it was observed that four of the top discriminators pertained to a duty area (Unit Administration) that appeared to be immaterial to distinguishing ODS/S from non-ODS/S experiences. Moreover, only two of these top 10 discriminators seemed to clearly favor warfighting experiences (i.e., "Supervise improvement of armor platoon vehicle fighting positions" and "Individually move under direct/indirect enemy fire"). With respect to the Armor AOAC study (Table 2), it was observed that all 10 top discriminators pertained to warfighting responsibilities. Still, the rank orders of the average TE ratings for non-ODS/S participants were more likely to be higher than lower in contrast to ODS/S participants. In particular, the rank order of the best discriminator (i.e., "Plan strongpoint defense at company/troop level") was relatively much higher for non-ODS/S raters (53.5 versus 183.5). Those tasks which had relatively higher rank orders of TE ratings by ODS/S raters included "Direct area reconnaissance at troop level", "Supervise treatment/handling of enemy prisoners of war (EPW)", and "Plan NBC training". Pertaining to MOS 93C, as shown in Table 3, only 4 of the top 10 discriminators reflected higher TE ratings by the senior NCOs who participated in ODS/S. Two of these tasks pertained to performing preventive maintenance checks and services (PMCS) on Air Traffic Control (ATC) equipment, (i.e., "Perform PMCS on AN/TSW-7A ATC Central)", "Perform PMCS on AN/TRN-30 (V) 2 beacon set (tactical/semifixed mode)"; one to unit administration (i.e., "Maintain/update reference material/miscellaneous instructions"; and one to ATC tower procedures (i.e., "Update Automated Terminal Information System (ATIS) information"). While performing PMCS on these items of equipment could well pertain to warfighting responsibilities, it could equally be stated that most of the tasks reflecting higher TE ratings by non-ODS/S participants could apply in a battlefield environment (i.e., "Prepare AN/TSC-61B flight coordinator central for movement", "Issue transponder modes/codes"). In short, the tasks which discriminated between ODS/S and non-ODS/S participants could not be explained easily simply by the average TE ratings for these complementary group of raters.

Conclusion and Implications. From these findings, it was clear that there was a remarkably high degree of reliability and correlation between raters who gained wartime fighting experience participating in ODS/S versus those who didn't who evaluated the Branch 12 and MOS 93C task inventories in terms of TE. Moreover, highly discriminating tasks which appeared to relate to warfighting responsibilities nonetheless were as likely to reflect relatively higher TE ratings by those supervisors/managers who had been involved in ODS/S as by those raters who had not. What this could suggest is that raters, based on their previous experiences and knowledge gathered in their military career, tend to evaluate tasks in terms of how critical they could be for mission success in a specific battlefield scenario. In particular, raters involved in ODS/S placed more emphasis on tasks related to desert offensive operations; on the other hand, non-ODS/S participants placed more emphasis on a wider range of operational scenarios. If similar results emerged from examination of other officer and enlisted studies conducted in these special surveys, then this would provide justification for the traditional method used by the AOSP in collecting TE information.

TABLE 1 - TASKS BEST DISCRIMINATING ODS/S RATERS vs NON-ODS/S RATERS
ARMOR OFFICER BASIC COURSE (AOBC)
(N=696 TASKS)

STEP ENTERED	TASK VARIABLE SELECTED	WILKS' LAMBDA	RANK OF AVERAGE TE RATINGS (ODS/S)	RANK OF AVERAGE TE RATINGS (NON-ODS/S)	DIFFERENCE (ODS/S vs. NON-ODS/S)
1	Perform PMCS on M113/M901-Series vehicle	.888	528.5	480.0	-48.5
2	Direct tactical road march at battalion/squadron level	.800	507.0	530.5	23.5
3	Review completed request/ authority for leave form (DA Form 31)	.703	385.5	331.0	-54.5
4	Supervise improvement of armor platoon vehicle fighting positions	.610	70.5	151.5	81.0
5	Respond to nuclear attack	.538	262.5	197.0	-65.5
6	Inspect unit mailroom	.448	498.5	508.5	10.0
7	Evaluate platoon training	.345	45.0	81.0	36.0
8	Individually move under direct/ indirect enemy fire	.273	203.0	257.0	54.0
9	Administer unit alcohol/drug abuse program	.233	398.0	372.0	-26.0
10	Determine/request data available from SIDPERS	.201	470.0	484.0	14.0

TABLE 2 - TASKS BEST DISCRIMINATING ODS/S RATERS vs NON-ODS/S RATERS
ARMOR OFFICER ADVANCED COURSE (AOAC)
(N=696 TASKS)

STEP ENTERED	TASK VARIABLE SELECTED	WILKS' LAMBDA	RANK OF AVERAGE TE RATINGS (ODS/S)	RANK OF AVERAGE TE RATINGS (NON-ODS/S)	DIFFERENCE (ODS/S vs. NON-ODS/S)
1	Plan strongpoint defense at company/troop level	.894	183.5	53.5	-130.0
2	Direct area reconnaissance at troop level	.741	94.5	149.5	55.0
3	Inspect/direct operator maintenance of 4.2" mortar ammunition	.627	540.0	508.5	-31.5
4	Supervise treatment/handling of enemy prisoners of war (EPW)	.534	483.5	567.0	83.5
5	Plan intelligence preparation of battlefield (IPB)	.465	52.0	19.5	-32.5
6	Plan tactical road march at batallion/squadron level	.410	67.5	95.0	27.5
7	Engage targets with M240 Machine gun from loader's station on M1/M1A1 tank	.370	521.0	446.5	-74.5
8	Plan military operations on urbanized terrain (MOUT) operation at company/troop level	.331	307.0	221.5	-85.5
9	Develop airborne assault plan	.301	567.5	557.0	-10.5
10	Plan/conduct nuclear/ biological/chemical (NBC) training	.272	97.0	159.5	62.5

TABLE 3 - TASKS BEST DISCRIMINATING ODS/S RATERS vs NON-ODS/S RATERS
MOS 93C SKILL LEVEL (SL) 1 COURSE
(N=480 TASKS)

STEP ENTERED	TASK VARIABLE SELECTED	WILKS' LAMBDA	RANK OF AVERAGE TE RATINGS (ODS/S)	RANK OF AVERAGE TE RATINGS (NON-ODS/S)	DIFFERENCE (ODS/S vs. NON-ODS/S)
1	Perform PMCS on AN/TSW-7A ATC central	.890	3.0	54.5	51.5
2	Prepare AN/TSC-61B flight coordination central for movement	.735	75.5	16.0	-59.5
3	Establish/maintain radar contact with aircraft within assigned area	.629	75.5	35.5	-40.0
4	Perform PMCS on AN/TRN-30 (V) 2 beacon set (tactical/ semifixed mode)	.535	65.0	145.0	80.0
5	Direct aircraft to test/range areas	.433	228.0	203.5	-24.5
6	Safeguard "For Official Use Only" (FOUO) material	.383	278.5	246.0	-32.5
7	Maintain/update reference material/miscellaneous instructions	.333	292.5	392.5	100.0
8	Update automated terminal information system (ATIS) information	.275	100.5	118.0	17.5
9	Issue transponder modes/codes	.232	231.0	215.5	-15.5
10	Receive/post/relay high density airspace control zones (HIDACZ)	.211	183.0	181.0	-2.0

THE ROLE OF TESTING IN ADVANCED INDUSTRIAL MANAGEMENT

Lawrence S. Buck
PRC Inc.

INTRODUCTION

Naval shipyards are being faced with a number of challenges to the manner in which they do business if not to their very existence. If the shipyards are to meet these challenges and to continue to do business at manning levels reasonably comparable to current levels, they will need to undergo some significant changes. A number of factors have emerged in recent years that will presage dramatic changes in the manner in which the shipyards conduct their business. The end of the Cold War has led to considerable debate concerning the size of the Navy with cries for downsizing the U. S. fleet. Cuts in the Defense budget have resulted in a 10% operating budget cut for the shipyards for the 1991-1995 period. Accommodating these cuts, in addition to keeping up with inflation, would require that shipyards achieve a 20% gain in productivity and efficiency. Other factors that will influence the manner in which shipyards will operate into the next century inclu‘.: increased competition for repair work from private shipyards, a reduction in complex ship construction, cutthroat pricing tactics of competitors, longer maintenance cycles of existing ships,and the need for resolution of existing organizational problems within the naval shipyards. The future of the shipyards and tens of thousands of shipyard workers will be determined by the choices and decisions made today with respect to the manner in which the shipyards will be managed and operate.

The Navy's response to the challenges faced by the shipyards is the Advanced Industrial Management (AIM) program. AIM is a comprehensive management program designed to improve the performance of naval shipyards by reducing the overall cost of ship maintenance while increasing effectiveness and efficiency. The net effect of AIM will be to dramatically change the manner in which naval shipyards conduct their business, well into the 21st. century.

The use of objective tests will be a contributing factor to the success of AIM. The objective of this paper is to discuss the role that testing will play in support of AIM. Specifically, tests will support and facilitate the Skill Designator (SD) program and the Shipyard Training Modernization Program (STMP), component programs of AIM. To this end, a discussion of AIM and its component programs is integral to an understanding of the role that testing will have in the overall picture.

ADVANCED INDUSTRIAL MANAGEMENT

The AIM program is an initiative sponsored by the Deputy Chief of Naval Operations for Logistics with the Naval Sea Systems Command (NAVSEA) as the executing agency. AIM will result in a comprehensive restructuring of the manner in which naval shipyards operate. The program will operate as a centerpiece program, pulling together and coordinating a variety of

existent and new programs. The basic tenets of AIM are listed below.

- Improve shipyard performance by improved work planning, estimating, and scheduling procedures.

- Reorganize the management structure to reflect a project organization with a single project superintendent, solely accountable for all aspects of a specific availability.

- Apply a zone technology approach to ship repairs.

- Improve data management and integration to increase the sharing of knowledge and planning products among shipyards.

- Standardize planning and work procedures and their products for use by all shipyards.

- Employ flexible work packaging by zone, trade skill, resources, etc.

- Create a single point of entry for each user to access automated information systems.

- Create a shipyard "corporation" by pooling resources and skills and consolidating information and training across shipyards.

- Provide workers with all work instructions, technical information, and all materials needed to perform the job at the time the job is assigned.

AIM is a dynamic program with the flexibility to incorporate successful shipyard programs while blending new programs into a cohesive package.

SKILL DESIGNATOR (SD) PROGRAM

One of the cornerstones of AIM is the SD program. In order to achieve an effective work force management tool, a method is needed to report actual planned and scheduled workload in a skill-based breakdown on which the work force can be based. That is, the actual scheduled work must be planned and estimated to specify labor skill requirements. Currently, work packages are planned and estimated for shop or shop/work center designators that do not correlate to specific labor skill requirements. For example, the welding shop might be designated as the lead shop for a job with pipefitters, riggers, and shipfitters listed as assist trades. Under the AIM and SD programs, the work would be assigned based on the requisite skills involved rather than on a shop or work center basis. In this respect, skill designators will function as an integrating tool for job assignments, human resource management, and personnel training functions within the shipyards in support of AIM objectives. SDs will be the driving force in effective personnel resource management.

SDs will be applied during the planning phase of an availability to facilitate workload and resource management. During the execution phase, SDs will serve as the parameters for individual work assignments. In this respect, work assignments will entail the participation of functional work team members for the completion of tasks that have previously been worked entirely within a shop or code. The functional team members will be determined by the respective SDs that they are certified for. New work processes and jobs will be assigned to the appropriate SD based on an analysis of the skills required for the job and the relationship of those skills to those described for a specific SD. Other benefits of the SDs include: establishment of a common language between naval shipyards for cost comparisons and for facilitating workload and work force comparisons, allowing for the exchange of planning products among the naval shipyards, and providing a basis for standardization of training plans and training requirements.

SDs have already been developed for close to 100 different task groups. Each SD carries a set of knowledge, skills, and abilities (KSAs) needed to perform the tasks encompassed by the SD. Two examples of SDs follow.

- Sheet Metal Layout:

  Measure sheet metal job at shipboard locations.
  Sketch (manually) sheet metal job/work in detail.
  Sketch (using CAD/CAM) sheet metal job work in detail.
  Use/operate numerical-control processes.
  Use/operate CAD/CAM processes.

- Generator/Motor Rewind

  Rewind armature.
  Rewind stator.
  Rewind field coil.
  Rewind transformer.

A central database is being constructed to include a central record of approved SD titles, knowledge and skill listings for each SD, and other supporting information. Associated data systems will contain information on the required training path(s) for an SD and courses available for each SD. The SD program will be aided by two existing shipyard programs, the Shipyard Skills Tracking System (SSTS) and the STMP.

OBJECTIVE TESTING IN SUPPORT OF AIM

Critical to the SD program is an ability to identify and/or certify expertise and experience in a particular SD or family of related SDs. One of the vehicles that will be used for this purpose is objective tests. The NAVSEA Industrial Skill Testing Center (NISTC), resident at PRC Inc. in Virginia Beach, will be responsible for the development and implementation of tests designed to verify SD related skills or to identify training needs related to a specific SD.

Over the past 6+ years, the NISTC has been deeply involved with the development of written and performance tests, at the journeyman level, for shipyard trades. During this time, tests have been developed, validated, and implemented for some 17 naval shipyard trades, encompassing around 85% of the shipyards' blue-collar work force. Tests for an additional eight trades are currently in the development or validation phase. The trade-skill tests will be used for a variety of purposes including: promotion of limited workers to journeyman positions, selection of external applicants for journeyman positions, evaluation of the apprentice program, evaluation of other training efforts, and determination of training needs in the incumbent journeyman work force. In order to facilitate the administrative burden of management of the trade-skill testing program, PRC developed an automated Test Processing System (TPS), which provides for item banking, automatic test generation, test scoring and item and test analyses, and production of a variety of statistical reports.

Extensive duty/task analyses were collected for the shipyard trades. These analyses include duty and task listings, generalized KSA statements called position requirements, and specific KSAs called item objectives. Each item in the item bank is linked to both generalized and specific KSAs as well as the duty that the item relates to. Coincidentally, yet fortuitously, the procedures followed in the development of the trade-skill tests, and the resulting data collected will serve the needs of the SD program well. Since each item is linked to a specific KSA, it will be a relatively simple matter to identify items from the item bank that address KSAs identified for an SD. The tests generated in this manner will be used to identify persons capable of working in an SD who have had no recent experience with the specific jobs encompassed. In addition, the tests will be used to certify competency of shipyard workers in a specific SD or SDs. The tests will also serve a diagnostic purpose in terms of identifying the areas of training required for workers to qualify for a particular SD. To this end, NISTC has been tasked with developing and issuing standard trade-skill testing materials to measure proficiency within an SD. The job analysis data associated with the trade-skill testing program will be merged with other job analysis data banks to provide the basis for development and assessment of an SD or to augment or validate existing SDs.

Under the Navy's Corporate Operations Strategy and Plan (COSP), the NISTC has been tasked with the development of progress tests for apprentices as well as special skill tests. Tests developed for these objectives will be indexed to SDs. Special skill tests will provide anchor point testing for AIM. Anchor points are defined as events within an overhaul that are important to the successful accomplishment of AIM. Included among these events are organizational or administrative aspects of the program and/or production tasks.

CONCLUSION

Changing political, economic, and business climates have raised serious challenges to the survival of the naval shipyards. The end of the Cold War, increased competition from private shipyards, and a shrinking U. S. Fleet have serious implications for the available workload for shipyards in the 1990's. If the shipyards are to survive without dramatic manpower cuts or eventual closures, they must start operating as more cost-effective and efficient businesses. Naval shipyards must improve their management practices and utilize their resources more efficiently. The Navy's AIM program is being implemented to provide shipyards with the means to survive through the remainder of the 90's and on into the 21st. century by reducing the overall cost of ship maintenance while increasing shipyards' effectiveness and efficiency.

AIM provides a focal point for coordination of a number of existent shipyard programs as well as new programs and approaches. The shipyards' trade-skill testing program will be an integral component of AIM and the SD program. Task analysis data collected for the trade-skill testing program will serve to verify existing SDs and to provide a basis for the development of new SDs. Tests will be developed to verify or certify skill levels in the work force that are related to specific SDs. Progress tests for apprentices and special skill tests developed for COSP will be used as anchor point tests for AIM. The development, validation, and implementation of some 17 trade-skill tests over the past six years has resulted in a solid base for the continuation and expansion of the use of objective tests in naval shipyards. The experiences gained through these efforts will stand the NISTC in good stead as we provide support to AIM and its goals and objectives for the naval shipyards.

Naval Officer Computer Utilization

LCDR Karen A. Doyle NC, USN
Bureau of Naval Personnel, Navy Occupational Development and Analysis Center (NODAC)

ABSTRACT

This study examines naval officers' involvement with all levels and types of automated resources. A Navy-wide officer survey was conducted by the Navy Occupational Development and Analysis Center in the second quarter of FY 91. The survey was mailed to a random sample of naval officers stratified across 35 officer communities and proportionated by the ranks of chief warrant officer to captain. For purposes of this study, the findings represent the sample. Descriptive statistics were used to interpret the data. The findings underscore the importance of computer literacy as an entry level skill for officers in all fields, as well as the need for standardization of software packages.

Background of Study

In April 1989, Director, Department of the Navy Information Resources Management (DONIRM) requested that NODAC conduct a Navy-wide occupational study to determine officers' involvement with all levels and types of automated resources. The intrinsic objective was to delineate who was using ADP resources, the functional areas being supported by ADP resources, as well as the relative time spent using these resources to achieve mission objectives. This paper presents a partial summary of the findings to this comprehensive study.

In response to this request, NODAC developed the Officer Computer Utilization (OCU) Survey with 52 "special interest questions" and 178 computer-related job-task statements. The special interest questions were provided by DONIRM staff and responses to NODAC's request for input from the fleet in June 1989. The job-task statements were developed from existing computer-related task lists used by the military services and inputs from Navy subject matter experts working in computer related fields.

The survey instrument was pretested to ensure its accuracy, completeness, and proper format. Norfolk, Virginia and Washington, DC were the two major sites selected to ensure the participation of officers across communities and in a variety of commands.

In July 1990, the survey was mailed to a random sample of 10,923 Navy officers (N) from an eligible population of approximately 61,000 (N). Efforts were taken to ensure representation from officers in the ranks of captain (CAPT/O-6) through chief warrant officer (CWO2/W-2) and from 35 Navy officer communities. A total of 6,764 (n) responses (62% of the surveys mailed) were returned. Descriptive statistics were used to interpret the data.

Findings

Who is using ADP resources?

Answering this question was accomplished by examining the responses to types of computer systems. Respondents were provided a list of seven types of computers: Microcomputers, Technical workstations, Minicomputers, Mainframes, Supercomputers, Integrated real-time systems, and Technical-tactical systems.

Respondents were asked "Thinking about YOUR JOB AS A WHOLE (not just the computer-related portion), how frequently do you use the following types of computers?" The following scale was provided for respondents to answer the question for each type of computer:

0 - Not Applicable (not available)
1 - Never (i.e., available but do not use)
2 - Almost Never
3 - Occasionally
4 - Routinely
5 - Frequently
6 - Almost Always
7 - Always

The respondents were divided into groups based on their involvement with each type of computer. A NO ACCESS group was comprised of respondents who answered Not Applicable (not available) (0). ACCESS NON-USERS were obtained from respondents who answered Never (i.e., available but do not use) (1). USERS were obtained from the respondents who answered Almost Never (2) to Always (7). Additionally, respondents who used at least one type of computer system were grouped into a category of ANYUSER (any computer user). The average level of use was determined by summing the responses from the USER group: Almost Never (2) to Always (7) for each type of computer.

Figure 1 presents a series of multiple bar graphs representing the percentage of responses from officers who used each type of computer and the percentage of ACCESS NON-USERS. At least one type of computer was used by 88% of the responding officers. The microcomputer had an average use of Frequently (5) by the majority of officers (84.2%). Approximately 20% of officers were using technical workstations, minicomputers, and mainframe computers, and 14% to 16% had access but did not use these systems. The average use of these computers was Routine. The supercomputer was used by only two percent of the officers, and on average, its use was Occasionally (3.2).

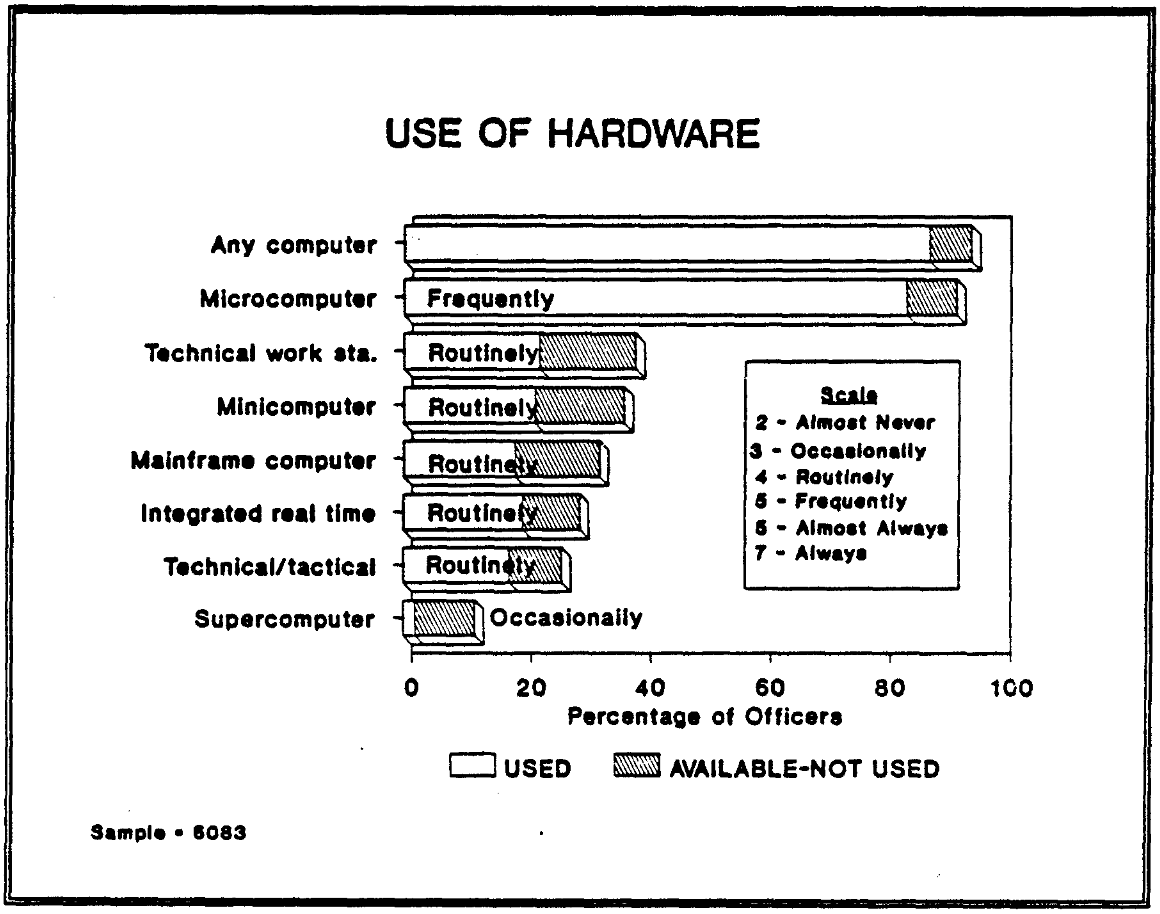


Figure 1. Percentage of Officers Who Used Different Types of Computers

What is the Relative Time Spent in ADP Efforts

Respondents were asked "On average, what PERCENTAGE of your work day involves computers, whether directly or indirectly?" Only the respondents who were involved with computers (i.e., used at least one type of computer or supervised at least one person who used a computer) were included in the analysis of this question.

On average, officers indicated that they spent approximately 31% of their work day involved with computers. Figure 2 displays the average percentage of work day involving computers for each rank. The average percentage ranged from a high of 40% for the ensigns, to a low of 21% for the captains.

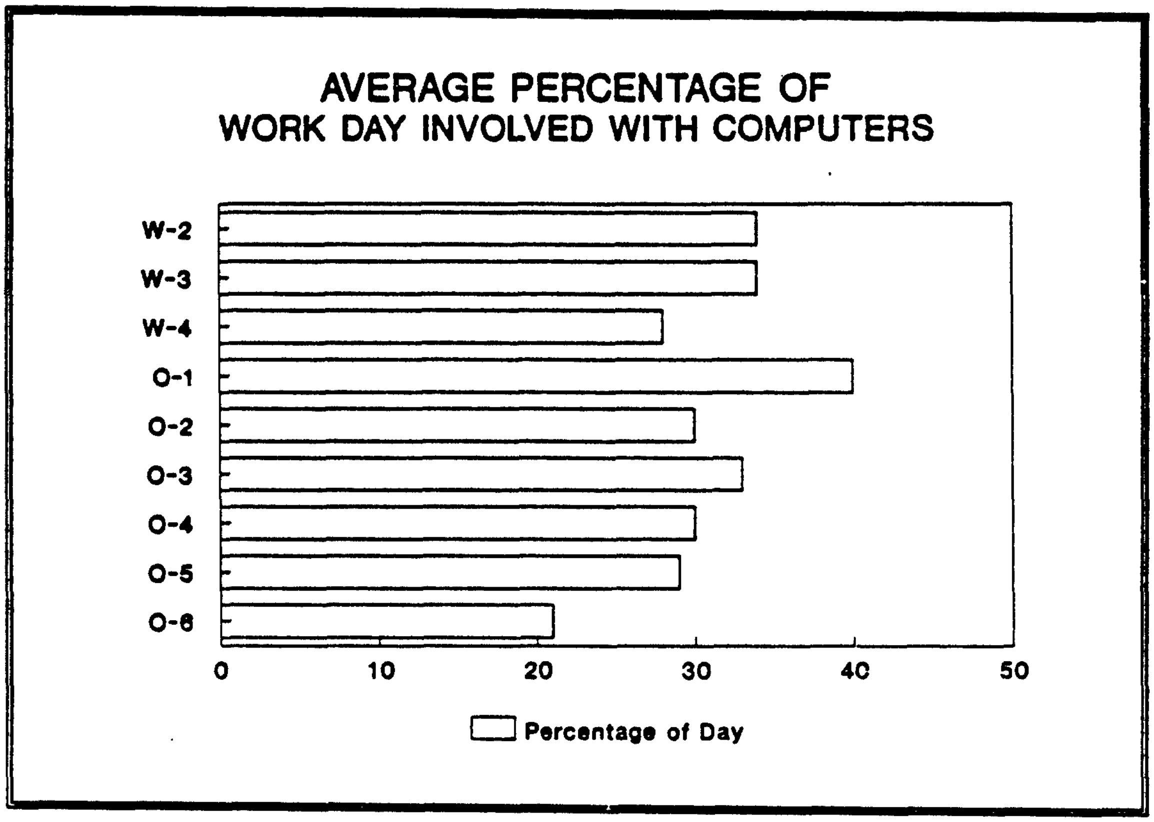


Figure 2. Average Percentage of Work Day Involved with Computers By Rank

<u>What Types of Software are Being Used?</u>

The survey provided a list of 27 commercial software packages and 13 word processing packages. Respondents were asked to indicate how frequently they used each brand, relative to their entire job using the following scale.

0 - Not Applicable (not available)
1 - Never (i.e., available but do not use)
2 - Almost Never
3 - Occasionally
4 - Routinely
5 - Frequently
6 - Almost Always
7 - Always

The commercial software packages were divided into eight sections: word processing, graphics, database, spreadsheets, interface, communications, statistics, and utilities packages. With the exception of the statistics packages, all were being

and utilities packages. With the exception of the statistics packages, all were being used on average Frequently (5) or Routinely (4). Figure 3 displays the percentages of computer users who employed various types of software and the percentage who had the type of package available but did not use it. The top brand chosen in each category is written on each bar in parentheses.

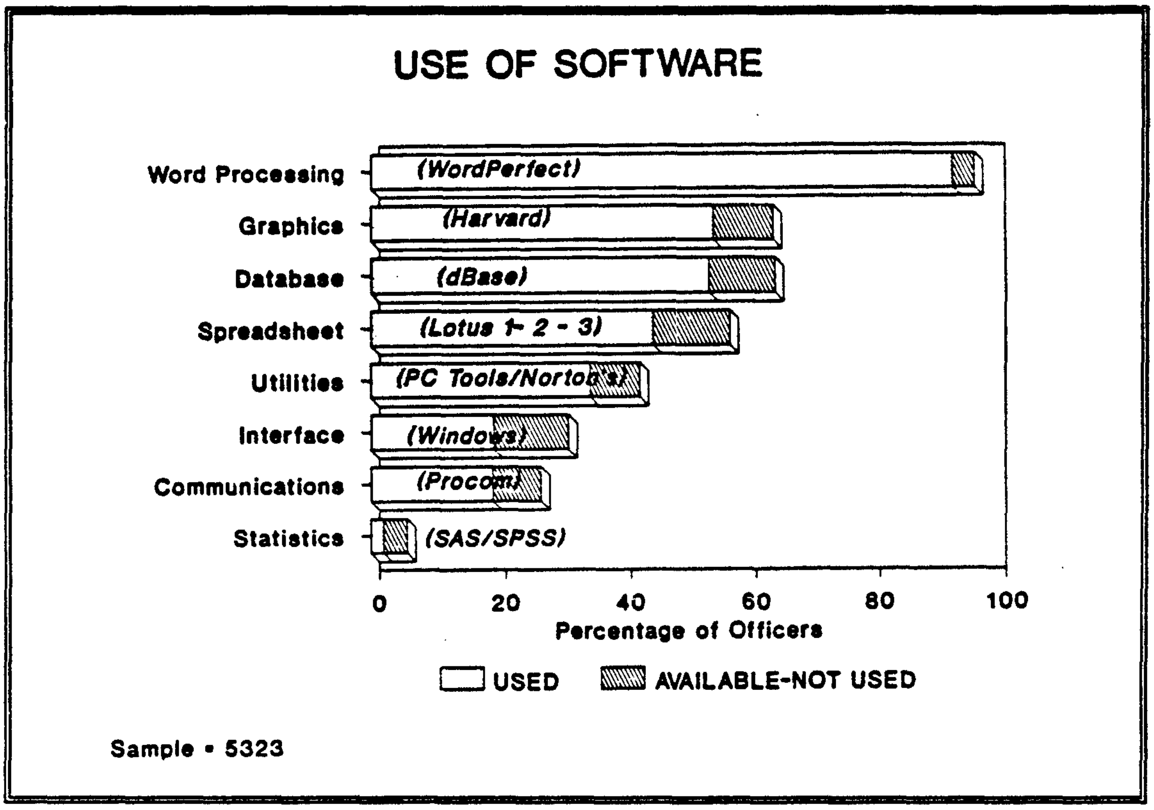


Figure 3. Percentage of Officers Who Used Different Types of Software Packages

The top four types of software packages used were word processing, graphics, data base management, and spreadsheets. At least two thirds of computer users had these available and at least half of computer users were using the top brand in each software type Frequently (5) or Routinely (4). Ten of the 13 word processing packages were used by less than 20% of the officers. *WordPerfect* and *WordStar* were chosen by more computer users than any other type of word processing software.

Harvard Graphics and *Desktop Publishing* were selected by more computer users than the *Enable Graphics* package. *Lotus 1-2-3* was the top brand of the spreadsheet packages. More respondents chose *dBase* than any other data base management package. Two thirds of computer users did not have communication or interface packages available.

Implications of the Findings

As written in the memo by Admiral Tobin to the officers chosen to complete the OCU, "Computer utilization at all levels within the Navy continues to grow exponentially." Of the officers represented in the survey, 88% used at least one type of computer system. At least 20% of all officers were involved with sophisticated computer systems beyond the level of the microcomputer. If not already a given, computer literacy beyond the level of word processing as well as knowledge of computer applications, will become a required entry level skill for all officers.

At least two thirds of computer users had word processing, graphics, database, and spreadsheet packages available. Half of these users also chose the top brand in each software category. This implies that the packages chosen meet the perceived needs of the users. It also suggests that a *de facto* standardization process is taking place. This standardization, if officially adopted, might contribute to increased productivity by decreasing retraining requirements and increasing system compatibility between commands.

The number of ACCESS NON-USERS in the various categories could be interpreted as being related to lack of training in the application of the hardware or software. However, the officer may not need the particular hardware or software application to perform his/her job, yet the command may require it to complete their mission. This finding may also relate to decisions regarding allocation of hardware and software assets.

The small numbers of officers who used communication packages are noteworthy. As the Navy puts greater reliance on electronic bulletin boards and automated systems to replace TAD conferences as a means of communication between levels and echelons of command, these software packages will increase in need and application.

Recommendations

- Evaluate the requirement for establishing computer literacy as an entry level skill for officers.

- Standardize software packages for major functional requirements.

- Conduct ongoing evaluation of computer asset allocation to optimize constrained resources.

- Evaluate the future requirements for communication software.

Using Event History Analysis to Study Task Data: An Update

Stanley D. Stephenson
Southwest Texas State University
Julia A. Stephenson
Kraft General Foods

Event history analysis is an analytical technique frequently used to study time-based data. It has the capability to handle binomial duration data (e.g., survival after cancer treatment) and also incomplete (censored) data. Event history analysis produces two main functions. The survival function computes the probability of survival through a time interval. The hazard function computes the probability of failure in a time interval, given survival up to the beginning of that interval. The hazard function has proven useful when comparing two groups. Event history analysis is now a well documented, frequently used statistical technique in many disciplines. Also, most popular statistical software packages now contain survival modules, attesting to the popularity of this technique (Goldstein et al., 1989).

In a study reported at the 1990 MTA Conference, (Stephenson & Stephenson, 1990), event history analysis was used to analyze the data of the type collected by the USAF occcupational survey program. Since actual survey task data were unavailable at that time, a data set was created. Task performance (yes/no) for a single, hypothetical task was modelled for 1000 airmen who were assumed to have entered the career field at the same time. Of the 1000 airmen, 300 (30%) were considered censored; i.e, these airmen left the career field before they had stopped performing the task in question. Of the 300 censored airmen, 200 were considered to have left the service at the 48th month which is the end of the first enlistment period and the point of heaviest attrition across all career fields. The remaining 100 censored data points were randomly distributed throughout the 13-72 month period of this study with the probability of censoring being higher prior to the 48th month.

Because the model was not based on actual survey data, the event history analysis functions could not be compared to existing data. However, the survival function was compared between the entire 1000 airmen data base and the data base without the censored data points included (n=700). The smaller data base would be typical of conventional analyses in which incomplete data (i.e., censored) are not known and therefore are excluded from the analysis.

The two survival curves were different with the greatest difference being at the 48th month, the point at which censoring is heaviest. At the 48th month, event history analysis simply 'knows' more than the conventional analysis techniques because it has more data (i.e., earlier censored data) to analyze. Event history analysis indicated that task life was longer than would be predicted from conventional analysis. After the 48th month, as the numbers in the two data bases become more equal, the two curves converge. Overall, the 1990 paper indicated that event history analysis might provide a different look at task life.

We subsequently obtained task performance data from the Jet Engine Mechanic, AFSC 426X2, career field, a career field which contains more than 5000 airmen. Once we examined the actual data base, we shifted from our earlier event history approach, which was based on a continuous data model, to a discrete data approach. There were two reasons for this change. First, the 426X2 monthly frequency counts varied greatly. Prior to investigating an actual survey data base, we had assumed that monthly frequency counts would be about the same in the first 48 months. However, monthly counts for the first 48 months for the 426X2 career field ranged from less than 20 to more than 100 airmen. This variability created problems in modeling monthly percent member performing differences. The impact of the monthly variation in frequency counts was magnified in later years. After the 15 year point there were fewer than 5 airmen in some months, again posing problems for modeling month-to-month differences.

The second problem encountered was task performance profiles. In our earlier study we had assumed that there were many tasks which had a high initial performance level, followed by a gradual but steady decrease in performance over the first few years of service. Such a profile does not exist within the 426X2 career field. Instead, we found three distinct task performance profiles. First, there was a start low/stay low profile. Event history analysis did not appear to be appropriate for these tasks both because this technique assumes either 100 percent performing or not performing at the start of the study and also because this analytical technique focuses on change. For start low/stay low tasks, change does not occur.

A second 426X2 task profile was a start high/stay high profile; this type of task might benefit from event history analysis. A third profile was a start at 0 percent/increase profile, a profile which might also benefit from event history analysis.

For these two reasons (monthly frequency variation and actual task profiles), a 20 year discrete Life Table method of computing the survival function was judged a more appropriate event history analysis model to use. Yearly percent members performing figures were computed by dividing the total number of airmen in each year into the sum of the number of airmen performing the task in that year. The Life Table estimates of the survival function are essentially the same as those produced by the product-limit estimates developed by Kaplan and Meier (1958) and used in our first analysis.

METHOD

We again started with a data base of 1000 airmen who were assumed to have entered the job (in this case, Jet Engine Mechanic) at the same point in time. Using the Life Table approach described by Lee (1980), we modeled the "Number Dying" figure, which in this case was the number of airmen who stopped performing a task in an interval, by computing the reported differences between adjacent time periods from the actual AFSC 426X2 data. E.g., if the difference in percent members performing between two adjacent intervals was five percent, five

percent was used to compute the number who stopped using the task during that interval. If the difference was in the opposite-from-predicted direction, a zero percent change was entered for that interval.

We modeled the Life Table "Number Lost to Follow-up" based on the year being studied. For years one through four, we assumed a 0.06 (6%) per year attrition rate. For years five through twenty, we assumed a 0.01 (1%) per year attrition rate. Not all attritors were considered censors. Those who left the career field during a particular year but who had stopped performing the task in question during an earlier year were treated as having 'died' (for task performance purposes) in an earlier year. Those who left the career field while still performing the task were treated as censored data. We computed censored data using both attrition figures and the percent of workers who were still performing a task in each year. For example, if we calculated that 50 airmen attrited from the career field during a particular year but that the survey percent member performing figure for the 426X2 career field for that year was 80%, we factored in 40 airmen (50 x .80) as being censored for that year. As required by the Life Table approach, the Number Exposed to Risk value was then reduced by half of the censored figure versus half of the attrition figure. Finally, we assumed both that 100 percent of the workers were performing the task in year one, the year in which the task is trained at a formal school, and also that the overall attrition rate was 50 percent after four years.

For comparison purposes, we also recorded actual percent member performing results from the 426X2 survey data. We treated all of the above data on a per year interval basis up to the 20 year point, the point when an airman can first retire. All airmen still performing the task at the 20 year point were considered censored data.

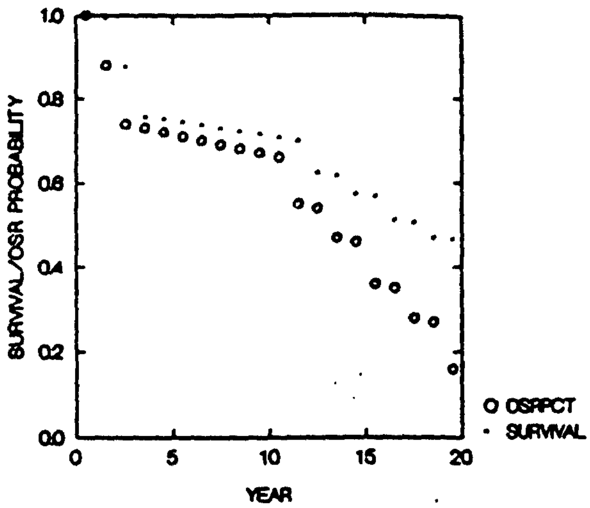
We performed the procedures described above on two tasks from the AFSC 426X2 task inventory. Task G406, "Remove or install engine plumbing" was an example of a start high/stay high task. Task B48, "Interpret policies, directives, and procedures," was an example of a start at 0 percent/increase task. For task B48, the event history analysis technique was reversed. That is, instead of 100 percent performing at the start of the time period (as was the case with task G406), 0 percent were performing at the start. Therefore, the analysis for task B48 was essentially "birth" analysis. This 'birth' analysis approach was appropriate because task B48 data are still binomial, even though the change analyzed was from non-performance to performance versus the other direction analyzed in task G406. For both tasks, comparisons were made between the survival function and the actual survey percent members performing figures for the 426X2 career field.

RESULTS

The survival function values and the actual 426X2 survey data values for task G406 are presented in Figure 1. Because of task G406's natural profile, the shape of the two curves Figure 1 are very similar until the 10th year. That is, after an initial drop, task performance maintains a plateau for about 10 years (start high/stay high). After the 10th year, the two curves

diverge primarily because the survival model contains information about censored data from the first 10 years.

Figure 1. Survival vs. OSR Data for Task G406



The interpretation of this graph is as follows. The OSRPCT (occupational survey report percent) plots are the actual data from the occupational survey results of the 426X2 career field; e.g., at the 20 year point, fewer than 20 percent of the survey respondents indicated that they were performing task G406. The SURVIVAL plots are the results of computing the survival function using the Life Table event history analysis approach. The 20 year survival function value of 0.48 is the probability that an airman who enters the career field at the beginning of his/her USAF career will still be performing task G406 in the 20th year. The survival function differs from the OSRPCT value in that it contains information from the previous 19 years while the OSRPCT does not; the OSRPCT value is specific to the 20th year alone. The survival function contains incomplete (censored) information from the many airmen who were performing the task early in their careers while the OSRPCT value is based on a small number of airmen who are still left in the USAF after 20 years. The two values obviously are providing different information. One can not be labeled better or more accurate than the other, although the OSRPCT data does come from actual survey results while the survival function does not. Nonetheless, it must be remembered that the OSRPCT data at the 20 year point is based on a much smaller number of airmen than is the OSRPCT data at the, for example, 10th year point. On the other hand, the survival function at the 20 year point has brought forward information from the much larger data base found in the earlier years. As compared to the OSRPCT values, the survival function values seem to indicate a much higher probability of task performance in the later years.

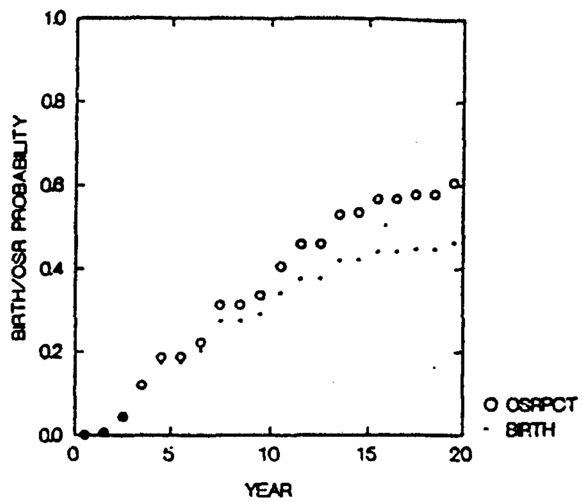
The general shape of task G406 is typical of other tasks in this career field in that once a task starts being performed, it

Secondary benefit has been the early identification of members with drug and alcohol problems. This

continues to be performed. Other tasks may be added to an airman's inventory, but beginning tasks do not quickly drop out of an airmen's inventory of tasks performed. Not until the worker reaches some degree of seniority does such a task's performance start to decrease. For task G406, this degree of seniority comes at about the 10 year point.

The results for task B48 are presented in Figure 2. B48 is a start at 0/increase task. As was the case with task G406, the two curves in Figure 4 are very similar in the early years. The later difference between the two curves is again due to censored data, only now the early years show non-performance; i.e., no one initially performs this supervisory task. Consequently, the censored data from the first years in service are non-performance and that is the partial information being carried forward by the survival/birth function. The accumulation of the large numbers of non-performing airmen causes the later years birth functions to indicate a lower probability of performance than does the survey data.

Figure 2. Birth vs. OSR Data for Task B48



DISCUSSION

For task G406, the survival function seems to indicate that the task has a higher probability of being performed in the later years than does the occupational survey data. Such an interpretation would suggest that proficiency in this task needs to be maintained longer than suggested by the survey data. For task B48, the birth function seems to indicate a lower performance in the later years than does the survey data. Such an interpretation would suggest that this supervisory task does not need to be trained as soon as would be indicated by the survey data.

Overall, the results of this study suggest that event history analysis might prove useful for investigating task life, in particular the life of a single task. As noted earlier, little

emphasis has been placed on the analysis of a single task, perhaps because of a lack of an appropriate technique. However, the conclusions from this study are derived from simulated data, although it should be noted that the modeling approach used in this study was conservative; i.e., this model does not exaggerate the differences between the survival function and the survey data.

It also appears that many, if not most, of the tasks (e.g., the start low/stay low tasks) in this career field would not benefit from more detailed analysis; percent members performing does not change over time for many tasks. However, the above statement does not take into account the fact that a start low/stay low task, as reported for the entire career field, might be a start high/stay high for one individual job type.

In summary, this study suggests that event history analysis might be appropriate for a small number of, perhaps important, tasks. But, the previous statement is based on an analysis of simulated data. To validate this statement, performance of an appropriate task (e.g., task G406) should be tracked over time, and the results compared to the modeled survivor function. A favorable comparison between this model and actual task performance measured over time would mean that this model could be used to add information (at relatively little cost) to the information currently available from the USAF occupational survey program.

REFERENCES

Goldstein, R., Andersson J., Ash A., Craig, B., Harrington, D., & Pagano, M. (1989). Survival analysis software on MS/PC-DOS computers. Journal of Applied Econometrics, 393-414.

Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. American Statistical Association Journal, 53, 457-481.

Lee, E. T. (1980). Statistical methods for survival data analysis. Belmont, CA: Lifetime Learning Publications.

Stephenson, S. D., & Stephenson, J. A. (1990). Using event history techniques to analyze task perishability: A simulation. Proceedings of the 32nd Annual Conference of the Military Testing Association, November 1990.

NAVY OCCUPATIONAL ANALYSIS IN A TOTAL FORCE ENVIRONMENT

Daryce L. Moore, LCDR, USN
Bureau of Naval Personnel Navy Occupational Development and Analysis Center

ABSTRACT

The elimination of the draft and adoption of the All Volunteer Force in 1973 caused the Department of Defense to reexamine its personnel strategy. In response, the Department of Defense (DOD) adopted the Total Force Plan. This plan recognized the fact that changing personnel requirements could no longer be met by adjustments in conscription quotas. Instead, flexibility and responsiveness to changing conditions would have to come from the combination of four categories of personnel support which comprise the total force: active duty, reserves, civilian fulltime employees, and civilian contractor support. Although the Navy Occupational Task Analysis Program was also inaugurated in 1973, its historical focus has been on task analysis for active duty personnel. Full implementation of the Total Force Plan, a necessity underscored by the changing world threat situation and the aftermath of Desert Shield/Desert Storm's precedent-setting use of reserves, requires that occupational analysis include tasks performed by both active and reserve forces.

BACKGROUND: THE NAVY APPROACH TO TOTAL FORCE

The Total Force Policy has evolved and changed as national security objectives have altered. When the Total Force Policy was adopted, military strategic planning rested on the assumption of a quick, decisive, potentially nuclear military campaign. During the next sixteen years, improved technology, the continuing Soviet military build up, and increasing realization of the horrors of a nuclear war prompted a change of planning assumptions. The focus shifted to preparation for a protracted, conventional war, with large ground troop mobilization.[1] This evolved into a strategy of partnership among all elements of the total force.

For much of the first two decades after adoption of the Total Force Plan, defense personnel requirements were generally met by active forces, civilian employees, and civilian contractors. Reserves played an important role through their contributions during drill and training periods, but the mechanisms for involuntary recall were not employed during any contingency operations during this period. An involuntary recall technically requires only the President's determination to do it, under Title 10 U.S.C. Section 673b, which gives the President discretionary authority to call up to 200,000 reservists for two successive 90-day periods. In practice, an involuntary recall sends a very strong message, both at home and abroad. During Operation Just Cause in Panama (1989) and Operation Earnest Will (escorting tankers in the Persian Gulf, 1987) reserves played a major but voluntary role.

---

[1]Report of the Reserve Forces Policy Board, DOD, Fiscal Year 1988, p. 89

Whether because national security requirements did not demand or because political conditions did not permit deployment of reserve assets involuntarily, the President did not exercise his 673b authority.

Navy force mix decisions during this period were admittedly influenced by a concern about "limited availability of Selected Reserve personnel" and "reluctance to initiate a reserve call-up."[2] As a result of these concerns, the Navy looked on its reserve assets as representing the difference between the active forces the country could afford to sustain during peacetime and the trained forces that would be needed at the outset of a conventional war.[3] The Navy had no plans to use reservists for short-term contingency operations.[4] This situation changed within days of Saddam Hussein's August 2, 1990 invasion of Kuwait.

Operation Desert Shield/Desert Storm changed the paradigm. It was a total force initiative, in which all elements comprising the total force played a significant role. This provided the first operational test of the workability of using the reserves to provide quick-response augmentation to active forces.

The aftermath of Desert Shield/Desert Storm, continuing budget crises at home, and the changing world threat situation all play a continuing role in the evolving military manpower strategy concerning the reserve forces. While the strategy of partnership remains unchanged, the goal of that partnership has changed from effective use of reserves in a full mobilization scenario to effective ongoing reserve utilization for a variety of conflict scenarios and mutual support for a smaller standing force.

HORIZONTAL INTEGRATION

The alteration in the desired reserve role had begun well before Desert Shield/Desert Storm. Former Secretary of the Navy John Lehman pointed out in a posture statement in 1986:

> Four years ago the Department of the Navy undertook a major reorganization of reserve components to move from a vertical to a horizontal relationship with the active forces. That means essentially that the reserves must provide immediate augmentation to the active force in time of emergency across the entire spectrum of warfare. It means also that, in peacetime, we rely on Selected Reserves to provide real-time fleet support across their mission areas.

Most experts see a direct relationship between what has come to be known as horizontal integration and training and readiness requirements. In order to train for a high state of readiness and facilitate integration when the reserves are activated,

---

[2] Total Force Policy Report to Congress, DOD, December 1990, p. 50

[3] A Report of the Navy's Total Force FY 91, DON, 1991, p. 3

[4] Total Force Policy Report, p. 49

reservists must have an opportunity to perform the tasks which will be required of them upon mobilization. The phrase "mutual support" was adopted to describe those training evolutions which also provide direct support to active duty units in fulfilling their missions. Examples include air logistics support for the continental U.S., air tanker services, fleet intelligence production, predeployment air combat refresher training, ship intermediate level maintenance, exercise support, cargo handling, construction support, chaplain and medical services and security group signal analysis.

ROLE OF TOTAL FORCE OCCUPATIONAL ANALYSIS

Despite the early shift to focus on horizontal integration and the overall success of reserve utilization during Desert Shield/Desert Storm, this first operational test of the policy revealed that the Navy still has a way to go in making the goal a reality. Occupational analysis provides one of the tools through which the Navy can monitor our progress toward horizontal integration and target areas of training required to facilitate the process. In August of 1991 the Navy Occupational Development and Analysis Center (NODAC) began developing a strategy to extend the Navy Occupational Task Analysis Program (NOTAP) to cover reserves.

The NOTAP process has been used since 1973 to identify tasks performed by members of all Navy ratings. The survey data generated by cyclical job task inventories administered to each rating form the basis of the Navy's published Occupational Standards (OCCSTDS), which establish minimum standards for personnel in each Navy rating. In turn, these are used in identifying training requirements, validating curricula, developing manning requirements, writing advancement exams, and making classification and force structure decisions.

By administering identical task inventories to active duty and Selected Reserve (SELRES) personnel, NODAC will establish the degree of similarity, thus horizontal integration, which exists between active and reserve personnel in various communities. In order to be smoothly integrated in a recall or mobilization scenario, reservists must be virtually interchangeable with their active counterparts in terms of ability to perform a given range of tasks. Identification of training requirements common to both elements is insufficient; training requirements can often be met in a classroom setting. Actual integration requires the ability to perform. The degree to which reservists have an opportunity to perform the same tasks as their active counterparts provides a "reality check" for the training system.

METHODOLOGY

By January 1992 NODAC had completed the initial feasibility research and developed the strategy described above. Consistent with the underlying premise of the Total Force Plan, a reserve officer was voluntarily recalled to active duty to serve as project manager to prepare the pilot study for administration. The study is being conducted in two phases, the first of which follows the standard NOTAP process.

The population identified for the pilot study was the Special Warfare (SPECWAR) SELRES community. The decision to start with officers rather than an

enlisted population came from the relative degree of need which existed. The use of OCCSTDS to prepare advancement exams which are identical for active and reserve components means that there is at least some degree of matching capability resident in the current system for enlisted personnel which is not subjective. The same cannot be said for officers. Studying the SPECWAR community specifically was an expedient choice. The active SPECWAR community survey was nearly complete and the reserve community's small size (71 SPECWAR reservists) meant that the pilot study could be managed closely and administered with minimum expense.

While task inventories must be identical between active and reserve participants, other aspects of the task inventory book must be tailored specifically to meet the needs of the individual component. While most of these differences are minor alterations in terminology for background variables (e.g., Reserve Unit Assignment Document instead of Officer Distribution Control Report), the scales used represented a major departure from active duty studies.

Liaison with commands responsible for various aspects of training and administration of reserves confirmed the preliminary assumption that relative time spent comparisons between active and reserve personnel would be useless and misleading. A reservist who spends two days a month and two weeks a year in training would not typically show the same pattern of time allocation in that status as an active duty member. The purpose of the instrument was to identify whether the reservist is performing the task at all, (i.e., a "do/don't do" approach), and if so, at what level (i.e., "do," "do and supervise," "supervise only"). To emphasize this aspect, questions for reservists are phrased in the past tense, while active duty questions use the present tense to connote continuity.

In addition, reservists may perform some tasks while fulfilling requirements for their mobilization billets, and others in a contributory support capacity. The difference is that mobilization billet requirements may be fulfilled during regular monthly drills or annual training with their active duty gaining command (i.e., the command to which they are scheduled to be assigned upon full mobilization). Other tasks are performed as contributory or mutual support for any active command which identifies a mission requirement. With the Navy's restructuring of the reserves subsequent to Desert Shield/Desert Storm, increasing emphasis is being placed on contributory support as a means of attaining training and horizontal integration. Correspondingly less is being placed on the traditional mobilization role. Therefore, a second scale for each task asked whether the task was performed in conjunction with the mobilization billet, contributory support, or both.

The survey was mailed to individual unit commanding officers via their local reserve centers. Initial responses have been slower than for active surveys, due in large measure to delivery time, since units drill only once a month.

Upon closeout, the data will be analyzed using ASCII CODAP to identify tasks performed by SPECWAR reserve officers. Commonality of tasks performed by active and reserve officers will be established on a "do" or "don't do" basis. In addition, the jobs performed by active and reserve officers will be identified. Tasks performed by active officers but not by reservists are a particularly crucial element for planners, and will be highlighted. Finally, NODAC will use task clustering as a method of determining whether there are patterns of difference between tasks association with

mob billet performance, compared to contributory support. The breakdown of tasks performed in contributory vice mobilization modes will be provided to total force planners for their incorporation into reserve restructuring decisions. Special Warfare Command and Special Operations Command, the two active commands with cognizance over most of the SPECWAR reserves will be provided the data for use in developing overall training requirements.

For purposes of prioritization, phase 2 involves a small, follow-on study. It will involve surveying subject matter experts, consisting of active duty SPECWAR commanding officers and executive officers, past and present. The instrument will consist of tasks from the active duty SPECWAR database. Specifically, it will include those with a performance emphasis (time-adjusted percent members performing) of 20 percent or greater. Participants will rate performance of the included tasks for level of importance as part of reserve training on a three-point scale: essential (1), necessary, but not essential (0), and not necessary (-1). The data will be analyzed using Lawshe's Content Validity Ratio[5]:

$$CVR = (n_e - N/2) / (N/2)$$

where

CVR = Content Validity Ratio

n_e = number of raters who rate the item as essential

N = number of raters

Given present time and resource constraints, this step is necessary to ensure that the tasks most important to horizontal integration receive top priority in creation and administration of training and as criteria to guide choices for annual training opportunities. A future possibility, if the pilot study accomplishes its objectives, is that this information may also be of use to members of promotion boards in comparing potential usefulness of members of the same competitive group.

SUMMARY

Ten years after the shift toward horizontal integration of active and reserve components began, the Navy has made great strides in making the goal a reality. Resource allocation during recent years has proven that the commitment to horizontal integration is real. However, the partnership aspects of horizontal integration need to be increased. Active duty commands need to regard the reserves as an integral and available element of the Navy's total assets, and understand how to optimize the contributions reserves can make. This understanding can be facilitated by total force occupational analysis, with its emphasis on comparability.

---

[5]Lawshe, C. H. A Quantitative Approach to Content Validity. Proceedings of Content Validity II Conf. Bowling Green State University, July 17-18, 1975, 24-37.

REFERENCES

A Report of the Navy's Total Force FY 91, DON, 1991.

Lawshe, C. H. A Quantitative Approach to Content Validity, Proceedings of Content Validity II Conf. Bowling Green State University, July 17-18, 1975, 24-37.

Report of the Reserve Forces Policy Board, DOD, Fiscal Year 1988.

Total Force Policy Report to Congress, DOD, December 1990, p. 50

**ATTRITION MANAGEMENT TOOL:
NAVY RECRUIT PSYCHOLOGICAL SCREENING**

by

**Imelda Ider
James A. Scaramozzino**

INTRODUCTION

In an environment where the Navy is experiencing end-strength draw-downs and constrained resourcing of programs, increased attention has been directed towards ensuring those accessed have the highest probability of being retained until they complete their enlistment contract. Using the principles of Total Quality Leadership, an aggressive review in FY-90 of first term attrition among Navy enlisted members indicated a loss rate of 34%. In numerical terms this equated to the premature loss of 26,500 sailors and an inefficient expenditure of $2.6 billion in recruiting and training dollars.

25% of these losses occurred during the first three months with the additional 15% following the fourth through the twelfth month of service. The remaining 60% of first term attrition occurred after arriving at the sailor's first permanent duty station/command. In that light, the decision was made to augment existing recruit screening processes. Initiatives and policies were executed to update literacy criteria, implement "Moment of Truth", revise the swim policy and align recruit physical fitness assessment procedures with the fleet's.

In FY-90, the trend of first enlistment losses started to change because Recruit Training commands were beginning to identify earlier those most at risk for discharge due to a mental disorder. In spite of increased attention, recruits most at risk for organizational delinquent behavior attrition were still being identified haphazardly and only after their level of dysfunction escalated. The scope of this paper addresses only the psychological screening process implemented to address thoses losses.

The Navy opted to implement the Air Force Medical Evaluation Test (AFMET) to ensure an equitable, quantifiable and consistent manner in early identification of those recruits most at risk for attrition as a result of a psychological condition. All three Navy Recruit Training Commands (RTC) (Great Lakes, Illinois; Orlando, Florida; San Diego, California) set in motion a standardized, three phase screening process, 1 October 1991. Its components include self-reported biographical data, standardized personality test results and clinical evaluation by a psychologist or psychiatrist.

AIR FORCE MEDICAL EVALUATION TEST (AFMET)

Since the mid 1970's, the Air Force has used AFMET to screen out those recruits likely to attrite from Basic Military Training (Fiedler 1990). It is a three phased process using: self-reports of life history variables, a preliminary mental health assessment, including completion of a standardized personality inventory and a mental health evaluation by a psychologist or psychiatrist. Recruits proceed from one phase to the next only if they meet a criterion indicating a high probability of need of further screening. Final disposition is based on the judgment of the licensed mental health professional. The interested reader is referred to Crawford's (1990) review of the history of AFMET.

Phase I- History Opinion Inventory (HOI)

The HOI is a 50-item, true-false self report of biographical history with a weighted total score range of 0 to 30. Higher scores indicate greater endorsement of problems prior to service (Fiedler, 1990). It measures eight categories: school and job problems; over concern with health, emotional instability; antisocial behavior, family dysfunction, withdrawn behavior; conflict with authority figures and immaturity. Validation studies are underway to re-standardize the HOI. The revision will have 70 items, two validity scales and remove some of the gender bias of the initial HOI (Fiedler, 1990).

Phase II-NEO Personality Inventory (NEO-PI) and Subjective Response Interview (SRI)

The NEO PI is a concise measure of the five major dimensions or domains of normal adult personality traits: Extraversion, Neuroticism, Agreeableness, Conscientiousness and Openness. NEO-Five-Factor Inventory (NEO-FFI) is a shortened version of Form S of the NEO-PI, consisting of five 12-item scales that measure each of the domains.

Subjective Response Interview (SRI)

The SRI is a structured mental status interview guide which collates pertinent clinical information with the patient's prior psychosocial history.

AFMET Phase III

Phase III involves a mental health professional (psychologist or psychiatrist) as the screener to

confirm or rule out the presence of any psychological pathology before any personnel action is taken. The Diagnostic and Statistical Manual criteria are used for this evaluation.

NAVY'S RECRUIT PSYCHOLOGICAL SCREENING PROCESS

In October of 1989 (FY-90), a combined line and medical initiative was begun to develop consistent, equitable and quantifiable procedures to identify recruits who were the most probable of attriting during their first term of enlistment for psychological disorders. Chief of Naval Personnel directed the process to be based on the Air Force's AFMET program. It was to begin after Moment of Truth; no gear was to be issued until the member completed the entire screening process; the evaluation was to be completed within three days of arrival at boot camp; the diagnosis of a mental disorder by a clinical psychologist or psychiatrist would result in mandatory separation for not meeting enlistment standards; if there was any doubt the individual would be given a trial of duty; motivated recruits who were in crisis would be provided up to three sessions of group therapy.

FINDINGS

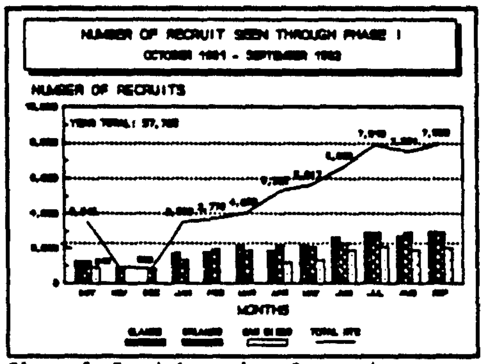


Figure 1: Total Accessions Screened

Figure 1 displays the individual RTC monthly recruit population who completed Phase I of N-AFMET and their cumulative totals (57,702 as of 30 September 1992). San Diego and Great Lakes were phased down in early FY-92 because of down-sizing initiatives. Great Lakes resumed processing recruits in January 1992. San Diego started up in April 1992. Total accessions for FY-93 is anticipated to be approximately the same.

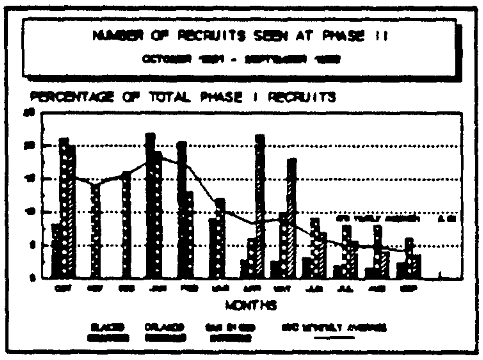


Figure 2 reflects the percentage of Phase I recruits referred to Phase II. In addition, it displays monthly trends experienced by individual RTCs and in the aggregate. For example, of the 2,683 Great Lakes recruits that went through Phase I during the month of August, 42 or 1.6 percent of Phase I went to Phase II.

Trends in these data are sensitive to two key factors: each RTC services a recruit population unique to the technical schools collocated with the RTC and the time of the year mirrors the quality of accessions. Consistency of throughput among the three N-AFMET units will never be completely parallel even though significant standardization in procedures was effected among the three N-AFMET units.

Recruits from Great Lakes normally stay for follow-on technical training in surface warfare courses. Orlando's schools train the submarine and the female communities. San Diego's recruits go to messmen, radiomen, internal communications, machinery repair and store keeper schools.

In the aggregate and individually the RTCs experienced a high Phase II throughput in the Fall, indicative of the low quality of accessions, and low throughput in the Summer, mirroring the higher quality of accessions.

The variation in monthly throughput was also attributed to implementation difficulties at each site due to unique environment and staff variables. By the end of the third month of implementation, the percentage of those referred to Phase II for each site progressively diminished. This is explained by the N-AFMET enlisted psychiatric technicians becoming more familiar with the screening procedures, instruments and criteria and the hardware and software difficulties were resolved.

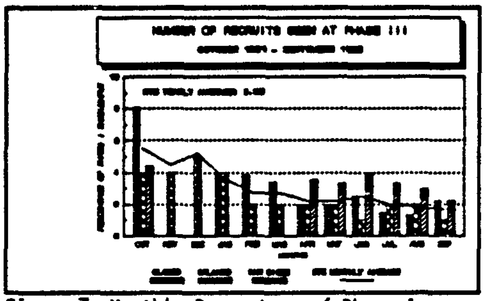


Figure 3: Monthly Percentage of Phase I Recruits Referred to Phase III

Figure 3 reflects the percentage of Phase I recruits referred to Phase III. In addition, it displays monthly trends experienced by individual RTCs and in the aggregate. For example, of the 2,683 Great Lakes recruits that went through Phase I during the month of August, 35 or 1.3 percent of Phase I went to Phase III.

The low percentage of recruits referred to Phase III reflects the value of the screening instruments, the skill of the enlisted staff to refer to the clinician only those with the highest probability to have a psychological condition and the time of the year which influences the quality of accessions. Variation among the three N-AFMET Phase III groups is again attributed to difference in the technical qualifications and educational background of the populations trained at each RTC.

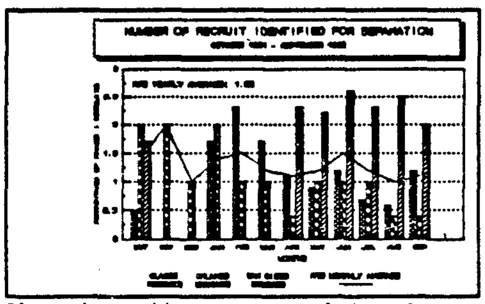


Figure 4: Monthly Percentage of Phase I Recruits Identified for Separation

Figure 4 reflects the monthly percentage of recruits identified for separation by individual RTC and the overall average. The criterion for separation of these recruits is bases whether or not they are diagnosed as having a psychological disorder by the clinical psychologist in-charge of the N-AFMET unit.

To ensure consistency in procedures among the three N-AFMET units and application of the DMS-III diagnostic criteria, considerable time was spent training the professional and technician staff. Additionally, N-AFMET referral and separation trends were closely monitored by headquarters staff.

Generally, the aggregate trend illustrated accession market traits. In the early months of the fiscal year the loower quality recruit results in higher numbers of separations. As the year goes on, accession quality improves (higher ASVAB scores, high school graududates) separations decreased.

RTC San Diego experienced the highest separation rate (2.5%) with the longest duration (five months). Separation rates at Orlando and Great Lakes were .7 percent and .6 percent respectively. The difference in these statistics is attributed to inconsistencies in Moment of Truth and N-AFMET procedures between San Diego and the other two RTCs. These inconsistencies have been resolved.

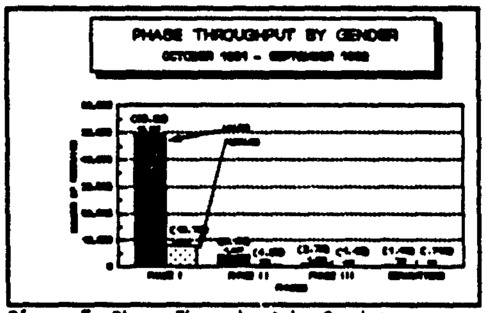


Figure 5: Phase Throughput by Gender

During the early stages of implementing N-AFMET, concern was raised about whether the screening process would be biased against females. Figure 5 reflects the male/female throughput for each phase of N-AFMET. Female recruits are only processed at Orlando Recruit Training Command. From October 1991 through September 1992, 4.8% of the 7,895 females processed were referred to Phase II. This is in comparison to the 9.2% of 49,807 males processed during the same period. The percentage of females referred to Phase III was 1.4% in comparison to 2.7% for males. Recruit females separated were 59 (.7% of the total female population) as compared to male recruit separations of 704 (1.4% of total male population). Revisions to the NOI mentioned early have been initiated to correct for the instrument's bias to not identify female recruits who have diagnoseable psychological conditions.

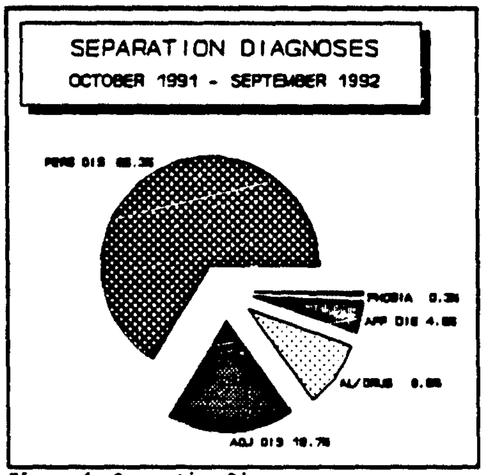


Figure 6: Separation Diagnoses

Figure 6 describes the various diagnoses for which recruits were separated during N-AFMET. The most prevalent diagnosis was some manifestation of a Personality Disorder (66.3%). The secord most prevalent diagnosis was for Adjustment Disorder (18.7%). A significant proportion of the recruits were diagnosed as Alcohol or Drug Dependent (9.8%). Phobias (.3%), Affective Disorders (4.8%).

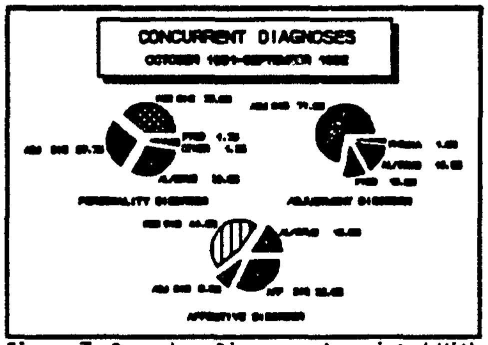


Figure 7: Secondary Diagnoses Associated With Seperations

Figure 7 describes the concurrent diagnoses of recruits discharged for a psychological disorder. Among the 66.3% of Personality Disorders, 39% had no other diagnosable condition; 30.3% were alcohol or drug abusers; 27.7% met the criteria for an Adjustment Disorder; 1.7% were diagnosed with Post Traumatic Stress Disorder (PTSD); and remaining 1.3%, classified as Other, included learning disabilities, phobias and eating disorders.

Among the 18.7% of Adjustment Disorders, 71.9% had no other diagnosable condition; 15.6% were alcohol or drug abusers; 10.9% met the criteria for PTSD; and 1.6% were phobic.

Among the 4.8% of Affective Disorders, 32% had no other diagnosable condition; 16% were Alcohol or Drug Abusers; 16% were Personality Disorders; 8% met the criteria for an Adjustment Disorder.

36% of the 713 recruits separated had some variation of an alcohol or drug related diagnosis. The early identification and separation of this population is anticipated to represent a cost avoidance of thousands of dollars as manifested in reduced safety accidents, reported alcohol/drug related incidents, suicide gestures, domestic violence, legal actions and management's time in attending to disruptive behavior.

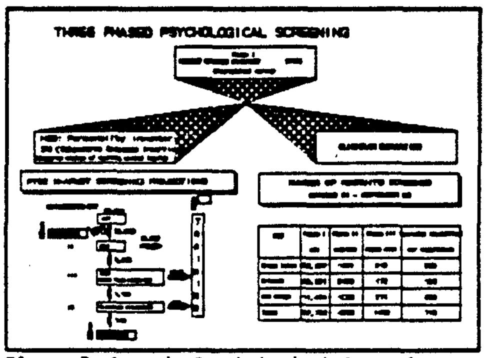


Figure 8: Recruit Psychological Screening Overview

Figure 8 summarizes the three phases of N-AFMET, the projected numbers for FY-92 to be screened and discharged and the actual numbers through the end of September 1992. Projections and actual numbers of referrals and separartions were very close.

DISCUSSION

Start-up cost was approximately $90K, which included the initial purchasing of equipment, training of staff and the hiring of a civilian analyst. Ongoing cost is approximately $45K: civilian salary and ongoing staff training. From a cost avoidance perspective the sunk cost was recovered within the first week of implementation. (It costs approximately $10K for a recruit to complete boot camp). Not issuing uniforms nor providing immunizations to recruits who were diagnosed within the first three days of training has resulted in cost savings of $100,000.

Secondary benefit has been the early identification of members with drug and alcohol problems. This population historically accounts for a large proportion of domestic violence cases and misconduct discharges.

Psychological attrition after the first week of during boot camp has significantly changed. This was anticipated because the population being identified with psychological disorders targets those who would attrite later in their enlistment.

Anecdotally, a change can be seen at the boot camps. There is more recruit company cohesion and recruits are acting out less frequently during boot camp and technical training. The numbers of recruits referred to the Recruit Evaluation Units after the screening has been reduced to half of what it used to be. These changes are the precursors of the positive aspects of the screening program.

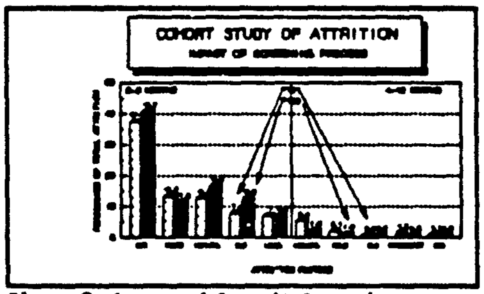


Figure 9: Impact of Recruit Screening on Attrition

Statistically, NAFMET has had a very positive effect on attrition. Figure 8 pertains. In a study of an FY91 and FY92 cohorts, attrition due to erroneous, fraudulent and entry level separation increase for the FY92 cohort during the first three months of enlistment. This attrition reflects attrition processed through the NAFMET screening. However, in the later months of enlistment, the impact of the screening is reflected in the significant decrease in attrition due to medical, drug, and misconduct reasons. Medical attrition was reduced by half, 5.4 percent in the FY91 cohort to 2.3 percent in the FY92 cohort. The recruits separated were identified earlier and in an organized manner saving training dollars.

REFERENCES

Crawford, L. (1990). Development and Current Status of USAF Mental Health Screening. Manuscript submitted for publication.

Fiedler, E.R. The Air Force medical evaluation test, basic military training and character of separation, unpublished paper, 1990, Lackland Air Force Base, TX, Wilford Hall Medical Center.

Characteristics of VSI/SSB Takers: Issues for Testing[1]

Kenneth A. Martell
and
Jane M. Arabian
Office of Assistant Secretary of Defense
(Force Management and Personnel)

From the post-Vietnam peak in FY 1987 to the completion of the planned drawdown in FY 1997, total active military end strength will be reduced 25 percent. By the end of FY 1993, the Department of Defense (DoD) will have reduced active duty military end strength by over 400,000--and will have accomplished three quarters of the planned drawdown.

The National Defense Authorization Act for FY 1992 and FY 1993 authorized separation incentives for officers and enlisted members who volunteer to leave the military, have at least six years of active service, and meet certain criteria established by each service. The authorized Voluntary Separation Incentive (VSI) and the Special Separation Benefit (SSB), which are, respectively, an annuity and a lump-sum payment, are temporary measures designed to help the Services reduce involuntary separations, align inventories with requirements, and permit programming of accessions to the proper sustainment levels.

The purpose of this paper is to describe the demographic characteristics of service members who have applied for either the VSI or SSB, focusing on the "quality" aspects. Fears have been expressed that these voluntary separation programs would induce quality service members to leave before those of lower quality. The standard index of "quality" is a score in the top fiftieth percentile (Category I-IIIA) on the Armed Forces Qualification Test (AFQT), an index of general cognitive ability used for enlistment, along with a high school diploma at the time of enlistment. While AFQT scores are a valid predictor of first term performance, their validity as an index of career performance, such as that of our target population, is still being investigated.

Therefore, it is reasonable to consider an alternative indicator of quality for these mid-career and senior career service members.

---

1 The views expressed in this paper are those of the authors and do not necessarily reflect the view of the Department of Defense. The authors wish to thank Ms. Valerie Franco at the Defense Manpower Data Center-West for creating the databases and programs used in our analyses.

Individual promotion timing is a logical alternative. Individuals promoted faster than the average for their service, grade, and specialty may be considered "high quality." The promotion timing of enlisted members within each Service is the direct result of comparisons made from results on several performance indices such as specialty test results, individual supervisor performance ratings, results of physical fitness tests, and other measures. Thus, service members who score better or are ranked higher than peers would be expected to be promoted at a relatively faster rate. This variable, however, has limitations as a cross-occupation, cross-service measure which will be discussed.

METHODOLOGY

The Defense Manpower Data Center (DMDC) established and maintains a VSI/SSB data base on each Service's applications for the voluntary incentives to include the applicant's Social Security Number (SSN), name, date of application, and other specifics. Matching applicants' SSNs with their individual master file data provides comprehensive, accurate information on each individual. More importantly, however, it allows analysis of the promotion rate of individuals compared to peers within each service, specialty, and grade. Individual records are used only for analysis purposes; privacy is preserved by reporting only statistical aggregates.

At the close of the FY 1992 VSI/SSB Program in August 1992, 51,801 applications for separation in FY 1992 had been received and 12,800 (Air Force only) for separation in FY 1993. A lag in the reporting of applications, especially from the Army, has resulted in approximately 11,000 fewer applications in the DMDC VSI/SSB data base used for this analysis. The sample 53,230 total DoD applications, or about 82 percent of the final FY 1992 program, is considered large enough to conduct demographic and promotion analysis and to develop accurate conclusions about promotion rates and quality for the overall VSI/SSB applicants in the FY 1992 program.

In order to make our exploratory analyses more tractable, we identified a sample of occupations opened to the VSI/SSB Program. Since the Services were given wide latitude, within the established law, to define their eligible population for force structure management purposes, it was difficult to make an a priori random selection of occupations to examine more closely. Therefore, we elected to select specialties with large numbers of takers (at least two hundred) across as wide a variety of specialties as possible (administrative, technical, combat, etc.). None of the Marine Corps specialties met our sample size constraints and are therefore not included in analyses.

The takers were set aside from each selected specialty to form the taker samples and the arithmetic mean time, in months-to-promotion to E-5 and E-6, and standard deviations were determined by specialty and grade. Finally, a random sample of cases was drawn from the non-takers in order to match sample sizes with the takers and permit t-test comparisons of taker/non-taker samples. The null hypothesis we are testing is that the mean promotion rate of the VSI/SSB takers is the same as the mean of the promotion rate of the non-takers. If true, the quality of the takers could be considered equal to the non-takers.

RESULTS AND DISCUSSION

Table 1 shows the percent of VSI/SSB takers by gender and race across all Services along with the DoD (inventory) proportions. The percentage of those self-selecting to leave the military is fairly comparable to the representation in the inventory. The greatest difference is exhibited by Blacks; they represent 29 percent of the takers but only 23 percent of the military population.

TABLE 1
VSI/SSB Program Demographic Analysis
DoD Inventory Versus Takers

		Inventory (%)	Takers (%)
Gender	Female	11	12
	Male	89	88
Race	Black	23	29
	White	70	65
	Other	6	6

The proportion, by grade, of takers versus the inventory is depicted in Figure 1. The greatest proportion of takers is at the E-5 and E-6 grade levels, corresponding to the mid-careerists. This is the population targeted most strongly by the program and supports our choice of looking more closely at these grade levels.

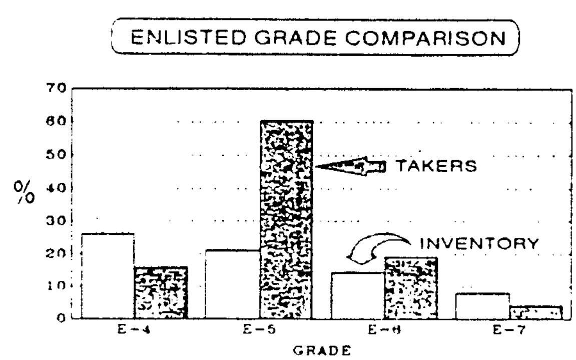


FIGURE 1.

With regard to "quality" indicators, Figures 2 and 3 show the aggregated proportions (across all Services) of AFQT category and promotion rates (average months to promotion), respectively.

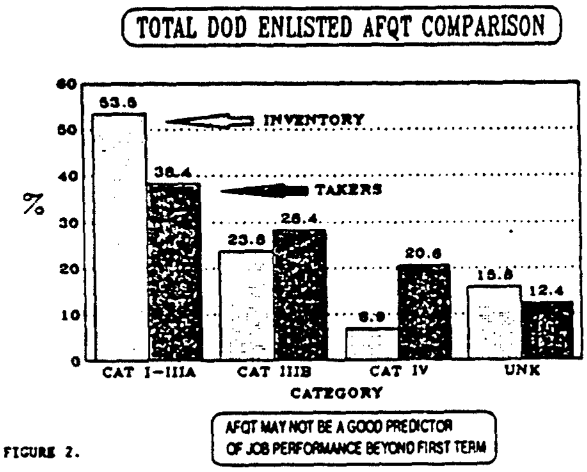


FIGURE 2.

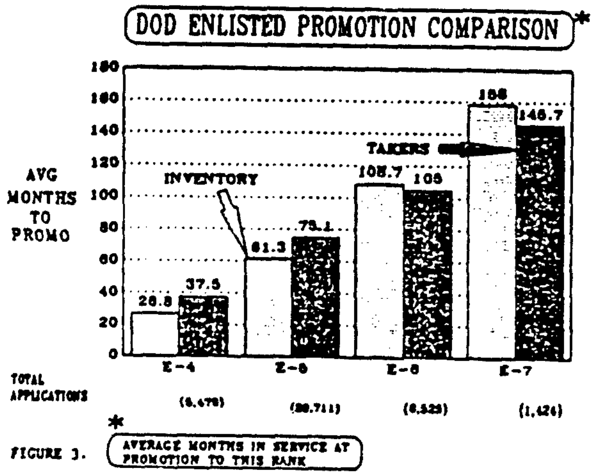


FIGURE 3.

The AFQT data suggest that overall, there is not a disproportionate quantity of higher quality service members (Category I-IIIA) self-selecting to leave the military (38.4 percent takers out of a population with 53.5 percent AFQT Category I-IIIA). The promotion rate data suggest that at the E-4 and E-5 levels, those with slower promotion rates (lower quality) are more likely to be leaving while at E-6 there appears to be little difference in promotion rates/quality among the inventory and takers and at E-7, those selecting VSI/SSB tended to be of higher quality than their peers.

The analyses by Service, grade (E-5 and E-6) and specialty support the notion that at the lower grade, E-5, quality is not being disproportionately lost (see Table 2). On the other hand, at least for the specialties we examined, we do not have evidence to support the notion that higher quality careerists are leaving (see Table 3). This may simply be a function of the specialties we happened to examine or an indication that promotion rate is not a viable index of relative quality at this grade level. At the higher grades, it is not unreasonable to assume that the system has already eliminated poorer job performers and that, at least with respect to promotion rates, the remaining careerists are equally good. Unfortunately, cell sizes were too small to allow analysis of potential differences among E-7.

TABLE 2 VSI/SSB PROGRAM PROMOTION TO E5

		NON-TAKERS			TAKERS				
SERVICE	SPEC	n	*M*	S.D.	n	*M*	S.D	t	Prob
ARMY	11H	66	51.59	17.52	66	71.71	29.75	-4.73	0.0001
	12B	224	58.05	23.40	223	65.34	25.32	-3.16	0.0017
	13B	326	53.62	27.33	327	68.52	25.83	-7.16	0.0001
	31C	189	54.48	22.30	193	55.04	23.05	-0.24	0.8077
	52D	197	47.42	17.41	198	58.10	22.18	-5.32	0.0001
	63T	123	58.13	20.75	129	58.81	22.49	-0.25	0.8020
	76Y	253	64.74	26.42	262	67.43	28.35	-1.11	0.2660
	94B	172	62.77	27.61	180	68.03	34.57	-1.58	0.1149
NAVY	AT	108	53.51	25.49	108	74.68	34.01	-5.17	0.0001
	MM	54	53.53	23.00	53	96.41	26.12	-6.90	0.0001
AIR FORCE	271	84	84.14	25.73	84	82.73	18.94	0.40	0.6876
	304	142	75.28	15.94	149	76.46	17.92	-0.55	0.5856
	452	552	77.90	20.15	552	81.43	19.46	-2.96	0.0031
	545	167	79.22	18.42	168	86.37	19.11	-3.48	0.0006
	645	659	77.97	21.38	649	82.10	18.48	-3.73	0.0002
	702	656	80.13	22.48	648	82.87	18.91	-2.38	0.0174
	811	1107	72.93	17.45	1115	79.16	18.18	-8.23	0.0001

TABLE 3 VSI/SSB PROGRAM PROMOTION TO E6

		NON-TAKERS			TAKERS				
SERVICE	SPEC	n	*M*	S.D.	n	*M*	S.D	t	Prob
ARMY	11H	89	91.42	26.55	75	98.26	21.9	-1.70	0.0919
	12B	197	88.49	22.42	200	92.13	24.09	-1.55	0.1209
	13B	270	83.75	25.23	281	84.22	21.33	0.24	0.8137
	31C	100	90.41	23.74	94	96.29	22.13	-1.79	0.0755
	52D	69	98.18	31.44	67	89.13	21.07	1.98	0.0503
	63T	79	94.56	25.14	82	92.63	24.88	0.49	0.6242
	76Y	403	91.82	22.46	406	94.00	22.18	-1.38	0.1666
	94B	167	100.04	27.68	169	101.04	24.7	-0.35	0.7288
NAVY	AT	89	101.07	26.37	88	107.27	29.12	-1.48	0.1401
	MM	105	81.94	25.05	107	103.79	25.05	-6.35	0.0000
AIR FORCE	271	24	153.58	24.14	24	147.95	21.76	0.85	0.4011
	304	INSUFFICIENT SAMPLE SIZE							
	452	127	138.18	26.66	137	137.84	25.74	0.11	0.9157
	545	INSUFFICIENT SAMPLE SIZE							
	645	112	145.3	26.67	116	140.93	26.31	1.24	0.2151
	702	101	149.11	28.39	109	139.23	24.28	2.70	0.0076
	811	174	140.72	27.16	178	135.00	27.36	1.96	0.0499

CONCLUSIONS

The Department of Defense is meeting its requirement to reduce the military end strength by approximately 406,000 between FY 1992 and FY 1997. The VSI/SSB voluntary incentive program is expected to result in 143,000 takers over that period or 35 percent of the total reduction. The grades of E-5 and E-6 make up 80 percent of the takers and have an average Year of Service of 11 years which is the

intended population for reduction. Demographic pictures for each Service are generally similar to that of DoD.

The original question concerning the quality of the VSI/SSB takers compared to the non-takers at this point can only be examined at the lowest level, that is, for each Service, grade and specialty. This is because each Service conducts its promotion boards, centralized or decentralized, differently. The added key dimension of specialties, with their inherent and unique qualification/promotion criteria, compounds the problem of aggregating the data across the enlisted force and generalizing to the overall Department of Defense. For an overall quality analysis of the career force using promotion rates, a common metric that somehow preserves the differences between Service policies would need to be developed.

Despite the difficulty of generalizing beyond specialty, grade and Service, the question of who is leaving and who is staying is still valid and warrants continued analysis. At present, our preliminary examination of promotion rate to each grade appears to be a satisfactory indicator of quality.

Preliminary demographic analysis using the current data base does show a slightly higher proportion of females compared to the inventory and a greater disparity between the Black enlisted inventory and the takers. The Services' selection of administrative specialties, with the higher proportion of Blacks and females, for personnel reductions to support the smaller baseline forces may be the most significant cause of these differences. It is important to note, however, that the long-term impact of the VSI/SSB program on females, minorities, and the quality of the force will not be as great as other personnel policies such as accession and retention requirements for each Service.

COMMANDER SURVIVABILITY ON THE NTC BATTLEFIELD

Dr. Robert F. Holz*
U.S. Army Research Institute

Background

Current Army doctrine (Department of the Army, FM 100-5, 1986) calls for bold, dynamic leadership on the high lethality battlefield envisioned by AirLand Battle. This requirement implies the risk of greater casualty rates for commanders. Given the increased lethality of current weapons systems and the requirements for commanders to "see the battlefield," assessments of commander survivability during rigorous training at the NTC takes on added importance.

Prior studies and analyses regarding the survivability of commanders during simulated training at the NTC (CALL, 1988; Doherty and Atwood, 1987), as well as the survivability of commanders during actual combat (Gal, 1985) suggest that a survivability rate of between 57% and 75% is likely.

The low level of engagements and low losses of U.S. personnel during Operation Desert Storm should not be viewed as the most likely scenario to confront U.S. forces in future battles. Rather, future battlefields may more likely approximate that envisioned in AirLand Battle Doctrine (to include low intensity conflicts) where the survivability of commanders will take on increasing importance.

The National Training Center.

The NTC has been designed as a realistic training ground for battalion task forces. Each battalion task force participates in about six force-on-force missions/battles during the two weeks it trains at the NTC. The force-on-force battles use the multiple instrumented laser engagement system (MILES) to record hits (and near-misses) on vehicles and players. These hits are electronically transmitted to computers at the NTC and form the basis for the data used in this report.

The NTC provides the best available laboratory for studying commander survivability on the AirLand Battlefield. Training is conducted under conditions that approximate, as close as possible, combat conditions and the instrumentation of weapons permits assessment of casualties.

Technical Approach

The sample used for the conduct of this study consisted of all deliberate attack and defend battles carried out by Armor and Mechanized Infantry Battalion Task Forces (TF) at the NTC during FY89. A total of 28 battalion task forces are represented with 73 battles constituting the sample. Thirty one of these battles were defend battles and 42 were attack battles.

Methodology

All data to be reported were obtained from the combat training center archive maintained by ARI at its Field Unit at the Presidio of Monterey. The data tapes were generated by the NTC Core Instrumentation Subsystem (CIS). Additionally, digital replays of selected battles were used to identify the tactical location of commanders whose vehicles became casualties during the deliberate attack battles. The data tapes permit one to identify the vehicle assigned to a commander by the presence of a unique three digit code (e.g., A66 would be the vehicle assigned to the Company Commander of Alpha Company) and to visually "track" that vehicle from start to end of a battle.

The findings that follow are based on the first time a commanders vehicle was reported lost to a direct fire kill during a given battle. Multiple losses and losses due to other factors (i.e. administrative, accidental or OC gun kills) were not addressed. As such, the data on commander survivability to follow reflects a more conservative estimate of such survivability than could be obtained had losses due to such factors as artillery or mines been included. Further, the data apply only to the vehicles of the commanders at the NTC and not to the leaders themselves. Therefore, regardless of whether the commander was in his vehicle when the vehicle was hit it was treated as an operational loss. Computations of commander survivability were calculated based on the proportion of commanders surviving a battle.

For each battle in the sample, tables were generated indicating changes in the status of the BLUFOR company commander's vehicles throughout the battle. The tables generated formed the basis for the results to be presented.

Results

Commander survivability & type of battle fought.

As can be seen in Table 1, the survival of company commanders across the 73 battles was 70%. Comparing the survival rates for these commanders in the attack and defend battles yielded nonstatistically significant differences. Survivability of company commanders does not appear to be related to the type of battle fought.

Table 1

Commander Survivability by Type of Battle

Attack	Defend	Combined
133/188 71%	96/139 69%	229/327 70%

Survivability and type of task force fighting the battle.

Survivability was found to differ for commanders in the Armor and Mechanized Infantry Task Forces (see Table 2). In the case of the Armor Task Forces, company commander survivability across the 38 battles fought was 63%

while their counterparts in the Mechanized Infantry Task Forces had a survival rate of 76% in the 35 battles fought (chi square=6.38, df=1, p<.05). Based on the above, it would appear that armor commanders are more likely to have lower survivability rates (at the NTC) then their mechanized infantry counterparts.

TABLE 2

Commander Survivability by Type of Task Force

Armor TF	Mech TF
96/152	133/175
63%	76%

Weapons firing by commanders who became casualties.

The extent to which a commander's vehicle becomes a casualty during a battle might be related, in part, to whether that commander's vehicle was firing its weapon(s) prior to being hit. To address this question, analyses of vehicle firings were computed for those company commander vehicles previously identified as having been hit by the OPFOR. Since it would be expected that the majority of commanders (and other BLUFOR players) would be firing their weapons during defend battles, the analyses were limited to those attack battles where one or more company commanders became casualties.

This procedure yielded a total of 30 attack battles in which 116 Company Commanders were engaged. Data from the Core Instrumentation System (CIS) was queried in order to determine whether, in a five minute period prior to becoming a casualty, these commanders were in turn firing their weapon(s) systems.

Results of this initial analysis indicated that across the 30 attack battles a total of 55 Company Commander vehicles were "hit" by the OPFOR. Of these 55 casualties, 15 (27%) were found to have been firing their weapon(s) systems in the five minute period prior to their becoming casualties themselves, see Table 3. Of these 15, 12 (80%) were found to be armor Company Commanders who fired their main gun on average three times in the five minute period prior to being hit.

TABLE 3
Commander Casualties and "Fighters"

	Armor TF	Mech TF
# "Hit"	36	19
# Firing Prior to "Hit"	**a** 10	**b** 5

a 8 out of 10 were Armor Cdrs
b 4 out of 5 were Armor Cdrs

Based on these analyses it would appear as though less than one third of all company commanders whose vehicles became casualties during attack battles were involved in fighting their vehicles prior to becoming casualties themselves.

Commander tactical location and survivability / mortality

Of greater importance than the issue of whether the commander was fighting his vehicle prior to becoming a casualty is the question of *where* that commander's vehicle was positioned when it became a casualty. Present doctrine states that commanders should position themselves forward on the battlefield so as to be able to influence the battle outcome. Accordingly an analysis of the tactical location of those commanders who became casualties and those who survived was conducted to determine whether location was associated with survivability / mortality.

To conduct this analysis battle replays were generated for the above noted 30 attack battles where one or more Company Commander's vehicle was reported as lost to OPFOR direct fire. The battle replays are graphical representations generated by computer and are based on data contained in the CIS data tapes from each NTC battle. These replays permit the analyst to identify individual players on the battlefield and to examine where they were located (in relation to both their own forces and the OPFOR) when they became casualties and their location at the end of the battle.

For each of the 30 battles, the individual vehicle assigned to a given Company Commander was identified at the start of the battle and then visually tracked through the battle up to the time when that vehicle was reported as having been killed by OPFOR direct fire or, in the case of surviving commanders, to the end of the battle. The tactical location of that vehicle vis a vis the OPFOR was then assessed in terms of distance (measured in kilometers) from the OPFOR forward line of defense. For example, if Alpha Company Commander's vehicle (A66) was previously identified as having been "hit" by the OPFOR at 0730 during a deliberate attack battle, then the A66 vehicle was located on the battle replay and its movement (on the NTC simulated battlefield) was tracked visually up to the time when it was reported to have been "hit." At that point, the tactical position of that vehicle could be assessed, e.g., A66 at 0730 for attack battle #1 was engaged in passing through the OPFOR minefields when it became a casualty. The same procedure was used for those commanders who survived the battle with their tactical location being noted at the end of the battle.

Of the 116 company commanders who fought in these 30 attack battles 55 of them became casualties and 61 survived. For the 55 company commander vehicles that were reported as having become casualties, 49 could be "tracked" by the battle replays. The movement of each vehicle, from the start of the battle until the time it was reported as "hit," was noted. To determine the tactical location of the commanders vehicle when it was hit, the battlefield was divided into three segments: Rear (over 6

km from the OPFOR barriers), Center (between 3 and 5 km from the OPFOR barriers), and Close (2 km or closer to the OPFOR barriers). A tally was made for each commander's vehicle (tank or infantry fighting vehicle) indicating where it was located when hit (see Table 4). For the 61 surviving commanders 51 could be "tracked" by the battle replays. The tactical location of each command vehicle was noted at the end of each battle.

Inspection of the data in Table 4 indicates that the vast majority (82%) of those commanders who were "hit" and three quarters (76%) of those commanders who survived were in the **forward or close-in** portion of the battlefield.

TABLE 4

Location of BLUFOR Commander Vehicles & Survivability / Mortality

	CLOSE	**CENTER**	**REAR**
% Cdr Vehicles Killed	**82%**	6%	12%
% Cdr Vehicles Survived	**76%**	12%	12%

These results indicate that an almost equal proportion of commanders who were killed and who survived had positioned themselves in the forward or close-in portion of the battlefield. Accordingly, tactical location does not appear to have been related to either survivability or mortality.

Discussion

The survival of commanders, during actual combat, is regarded as critical to battle outcome (CALL 88-1). The survival rates for the company commanders in the present sample give grounds for further thought. The survivability of commanders in prior U.S. combat indicates that "where the battle was of high intensity and of critical importance, a loss rate of roughly 30% among commanders during such battles is generally found" (Personal correspondence with the CAC Historian, 1991). The findings from the present study are in accord with these figures.

Data from the Israeli Defense Forces (Gal, 1985) on commander survivability, derived from the high intensity battles fought during the 1973 Arab Israeli War and the 1982 Lebanese war (wars that closely resemble those envisioned by AirLand Battle) reveal a loss rate for Israeli officers of 28% in the 1973 war and 25% in the 1982 conflict. The IDF attributes these leader losses to its policy of requiring leaders to lead from the front, risking themselves first, serving as an example to their men. The findings from the present analysis mirror those reported by the IDF with 30% of Company Commanders being lost during the conduct of the 73 battles fought.

The findings dealing with whether company commanders whose vehicles subsequently became casualties were fighting their vehicles prior

to being hit and those addressing the tactical location of these same commander vehicles when actually reported as hit indicate that current Army doctrine is being followed. In the first case, the majority of company commanders whose vehicles were hit were not found to have been personally fighting their vehicles. In the second case, the majority of the company commanders who survived and who became casualties were found to have been positioned in the forward area of the battlefield.

These findings point to the need for units to develop, and practice, commander succession during training given the realities of mortality on the battlefield.

---

* The views expressed in this paper are the authors and do not necessarily reflect those of the U.S. Army Research Institute or the Department of the Army.

REFERENCES

Commander Survivability: Lessons Learned 88-1: Center for Army Lessons Learned, U.S. Army Combined Arms Center, Fort Leavenworth, Kansas. 1988.

Department of the Army. F.M. 100-5. **Operations**. May, 1986.

Doherty, W.H. and Atwood, N.K. **Commander Survivability at the National Training Center.** BDM Corporation Unpublished Report. 1987.

Gal, R. **Leadership and Battle Behavior.** Army Trainer. 5, 6-11. 1985.

Morris, R. Command Historian, Combined Arms Command, Fort Leavenworth, Kansas. Commander survivability in U.S. battles, a historians perspective. Personal communication. 1991.

Using Simulation to Support Testing: Implications of a HARDMAN III Application

Laurel Allender
Army Research Laboratory
Human Research and Engineering Directorate
D. Michael McAnulty
and Carl Bierbaum
Anacapa Sciences, Inc.

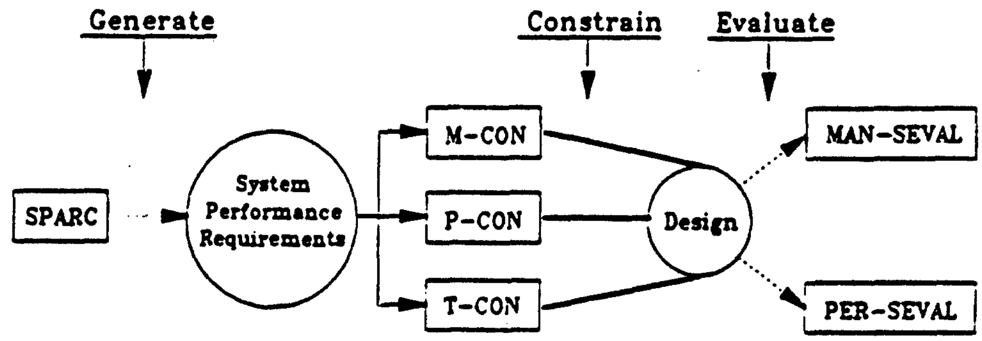
The downsizing of the Army has resulted in a change in the strategy for acquiring new systems. The emphasis is on building the technology base, on research and development--not on production and manufacture. Such a strategy assures a continued technological advantage without the acquisition of unneeded hardware. The question looms, however, what does such a strategy say about the Army's readiness? How can the Army, or any of the armed services, be prepared for conflict using new systems when those new systems are not fielded but only poised for production, when those systems lack the benefit of iterative field testing?

The answer has many aspects: interface design, soldier selection, reserve training, manufacturing capabilities, and, of interest here, simulation and modeling. Simulation and modeling are identified among the critical technologies for the future in the Army Technology Base Master Plan, Feb 92. According to the plan, simulation and modeling will be used for "testing of concepts and designs without building physical replicas."

Before proceeding with the discussion of an Army-developed simulation and modeling tool applied to an aviation problem and its implications for testing, one critical stipulation must be made: The soldier must be part of the solution. Soldier performance must be represented in simulation and modeling efforts. Without this representation, all of the historical arguments about why soldier considerations must be part of system acquisition which led to the Army's Manpower and Personnel Integration (MANPRINT) program will have been ignored. Suffice it to say, that now, more than ever, when resources for system acquisition, operation, and maintenance must be managed carefully, soldier considerations must be included and the tools for doing so must be exercised.

HARDMAN (Hardware vs. Manpower) III

HARDMAN III is a suite of six software modules that uses simulation and modeling methods to represent combined soldier-system performance. Further, it links performance to overall mission effectiveness. It was designed to support Army analysts and decision-makers early and throughout system acquisition and fielding.



SPARC: System Performance and RAM Criteria Estimation Aid
M-CON: Manpower Constraints Estimation Aid
P-CON: Personnel Constraints Estimation Aid
T-CON: Training Constraints Estimation Aid
MAN-SEVAL: Manpower-based System Evaluation Aid
PER-SEVAL: Personnel-based System Evaluation Aid

Figure 1. Six HARDMAN III modules in the context of system criteria, constraints, and evaluation.

The six HARDMAN III modules are depicted in Figure 1. The first module, the System Performance and RAM (Reliability, Availability, and Maintainability) Criteria Estimation Aid (SPARC), is used to help set realistic system and mission performance criteria through the use of task network modeling and library data bases. The Manpower Constraints Estimation Aid (M-CON), Personnel Constraints Estimation Aid (P-CON), and Training Constraints Estimation Aid (T-CON) are used to identify the numbers of soldiers, their skills and abilities, and the training resources likely to be available in the fielding years that will present constraints on the system. Manpower and personnel projection models, performance data, and historical training data underlie these three constraints modules.

The last two modules, Manpower-based System Evaluation Aid (MAN-SEVAL) and Personnel-based System Evaluation Aid (PER-SEVAL), are used to evaluate system designs with respect to the manpower crew sizes and personnel characteristics required by those designs to achieve the system performance criterion. Both use task network modeling and draw upon libraries of mission and individual performance data. In addition, PER-SEVAL is used to evaluate performance under various conditions of environmental stress such as heat, continuous operations, and protective clothing. Each of the six modules can be used in a stand-alone mode or as part of an integrated analysis.

HARDMAN III was developed as a desk-top concept. The software runs on an IBM-compatible 286 PC or better, with M-S DOS 2.2 or higher, 640K random access memory, an enhanced color graphics display, and a minimum of a 20 MB hard drive. The modules were designed with a menu-driven interface, on-line help and data references, and extensive library data on Army systems and Military Occupational Specialties (MOSs).

Forward Arming and Refueling Point (FARP) Turnaround Time

The problem of the FARP turnaround time was identified as a candidate for a HARDMAN III analysis by the US Army Aviation Center and School, Fort Rucker, AL. The FARP is a temporary arming and refueling site for helicopters located forward in the battle area. The turnaround time requirement is threat-driven: Helicopters must be serviced and the FARP moved before the threat detects it and attacks. Currently, the Apache helicopter, the Army's premier attack helicopter, has a turnaround time of roughly 45 minutes. The threat requirement has been assessed at 15 minutes. The issue, then, is how current FARP operations and equipment can be modified or augmented so that the Apache (see Figure 2) can be rearmed with up to a full load of 1100 rounds of 30 mm ammunition, 38 rockets, and 8 Hellfire missiles, and filled with 200 gallons of fuel within 15 minutes.

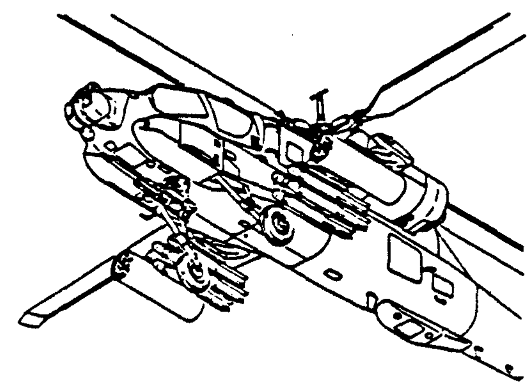


Figure 2. The AH-64 Apache helicopter weapon systems (viewed from below).

Analysis Method and Procedure

The HARDMAN III-FARP analysis was guided by the FARP organization described in FM 1-104, observation and videotaping of a sustainment training exercise, subject matter expert input, and an earlier report and videotape of an operational exercise (MTA, Inc., 1989). Also, information was reviewed about a proposed onboard sideloader (Western Design Corporation, 1988) intended to replace the existing frontloader for the 30 mm chain gun. Reloading the gun was identified by Fort Rucker as the most promising task for reducing turnaround time.

The baseline FARP model represents current, typical FARP operations. In the baseline, the 30 mm gun is reloaded using the frontloader which must be attached and detached as a part of the reload procedure. Two MOS 68J, aircraft armament/missile system repairers, have the responsibility for arming and one MOS 77F, petroleum specialist, has the responsibility for refueling. The conditions (environment, training, ammunition availability, etc.) are assumed to be optimal.

For the FARP analysis, MAN-SEVAL was the principal HARDMAN III module used. The steps in a MAN-SEVAL analysis are first to identify the mission and its performance criterion. Then the mission is decomposed into its constituent networks of functions and sub-functions (i.e., tasks). Functions and tasks can run in parallel paths, occur with some probability less than 1, or repeat some specified number of times. Each function and task also has an associated criterion. The critical performance data are

input at the task level: most likely time to perform and fastest time (from which are calculated the mean and standard deviation used in the model).

Once the mission is decomposed, the crew members are identified, their availability to perform a given task is assessed, and the tasks assigned accordingly. Next, the mission is executed from the bottom up, drawing on the distributions of performance times. Last, the resulting task, function, and mission times are compared with the criteria to determine overall mission success or failure.

The details of the building the FARP models will not be reviewed here but see McAnulty, Bierbaum, & Allender (1992) for a complete discussion. Since the HARDMAN III mission library is concentrated on weapon systems, no data were available to feed the FARP baseline. Instead, a preliminary model was built from scratch based on the observations, videotapes, and existing reports. After several iterations, the baseline was completed and verified against the source data.

Once the baseline was established, the modifications to simulate various options for achieving the 15-minute turnaround criterion were undertaken. Three options were evaluated singly and in combination--equipment, personnel, and logistics (see Table 1). The rationale for the equipment change is that, with the onboard sideloader, the 30 mm loading device becomes an integral part of the helicopter and does not need to be attached and detached every time the gun is loaded. Access to the onboard sideloader is also thought to be easier compared to frontloader access, which should also speed performance times. Although in the downsized Army, an increase in personnel is not readily justified, it is an option that demands consideration given the criterion. The logistics option of a partial reload is based on probable usage rates. Whereas a complete reload of either the missiles or rockets is likely to be expended during a typical Apache mission, the full complement of 1100 rounds of 30 mm ammunition is not. Therefore, a partial expenditure of 30 mm ammunition requires only a partial reload.

Making the modifications to the models was relatively straightforward. To change from the frontloader to the onboard sideloader, the task times for reload were reduced based on the proposed concept (Western Design Corporation, 1988). Also, the number of tasks in the "prepare gun for loading" function was reduced from 9 to 3. To add personnel was somewhat more complicated. In the two 68J condition, a task might be assigned to one 68J. In the three 68J condition, that same task might be shared by two 68Js. From the modeling point of view this meant that the task had to be duplicated and put on

Table 1. FARP Baseline and Improvement Options for Equipment, Personnel, and Logistics

	Baseline	Improvement Options
Equipment	Front-loader	Onboard Sideloader
Personnel	2 68J	3 68J
Logistics	Full reload of 30 mm (1100 rounds)	Half Reload of 30 mm (550 rounds)

a parallel path so that the two 68Js could work simultaneously. To reduce the amount of 30 mm ammunition to be loaded, the number of repetitions for that task were reduced. Each of the resulting models was run 5 times to represent the five Apaches typically found in a company.

Simulation Results

Introduction of each of the three options--equipment, personnel, and logistics--singly reduced the turnaround time compared to the baseline condition (see Table 2). Introduction of the onboard sideloader and the third 68J each reduced the turnaround time by nearly 30%, from 41.46 to 29.13 and 27.92 respectively. A half reload considered alone resulted in less than a 20% improvement, 33.35. Combining the equipment and personnel options further reduced the turnaround time to 18.63 and a combination of all three options reduced the turnaround to 15.78, a close approximation of the criterion. However, the amount of 30 mm ammunition had to be reduced to one-third of a load to meet the 15-minute turnaround time consistently.

Implications for Testing

As this effort was beginning, the plan was to use a scheduled field test of the sideloader option to validate the simulation results. The test, however, was slipped and the tables were turned. The results of this application of HARDMAN III to the FARP are being used to refine and guide the operational test planning of the onboard sideloader, now scheduled for spring 93. The test trials with the biggest expected payoff can be selected based on the simulation results.

The cost- and time-effectiveness of simulation and modeling is evident in other ways. The use of HARDMAN III cost 5 person-months including some time to learn to use the tool. Of that 5 months, nearly 4 months were devoted to planning, building, and validating the baseline, but only 1 to build and run all of the options. This compares extremely favorably to an estimated order of magnitude increase in man-months required for a full-scale operational testing effort.

Further, many more options can be tested in the simulation and modeling environment than is practicable in a field test. For example, several successive simulation runs were used to determine the maximum number of 30

Table 2. Average Turnaround Time in Minutes for the Three Options and Their Combination

	Equipment			
	Frontloader		Sideloader	
	Personnel		Personnel	
	2	3	2	3
Logistics				
Full Reload	41.56*	27.92	29.13	18.63
Half Reload	33.35	22.84	23.85	15.78
One-third Reload	-	-	-	14.94

* Baseline condition

mm rounds that could be loaded and still achieve a stable estimate of a 15-minute turnaround. These additional trials were obtained at virtually no cost when compared to the cost of running comparable field trials.

Some of the "cost" of doing the FARP analysis, building the baseline, is already being recovered. When a HARDMAN III model is built, it becomes part of the available library. Comanche analysts have stated a need to conduct a HARDMAN III analysis on the Comanche FARP, which also has a 15-minute turnaround requirement--at night in environmental protective gear. The baseline and the various options modeled here can be readily adapted for the Comanche analysis. Also, another modeling effort looking at the entire logistics arena, from factory to helicopter, is using the data and models built here as a component in a larger simulation. This analysis is currently serving as the corner piece for an integrated analysis using all six HARDMAN III modules.

Simulation also has the advantage of being entirely reproducible, and in the case of HARDMAN III, exportable to other users. The answers provided are quantitative and timely. Also in the case of HARDMAN III, the answers are readily accessible to the users. The software runs on a desk-top and requires no special programming background to use.

Sometimes in field tests, there are serendipitous discoveries of some aspect of performance. That is not precluded in the simulation and modeling environment. In this analysis, the task sequencing and sharing was optimized to minimize "dead-time." This performance optimization is being reviewed for its contribution to training of procedural tasks.

In summary, the use of simulation and modeling has implications for testing as a resource multiplier. The HARDMAN III application to the FARP problem was timely, cost-effective, quantitative, thorough, and captured serendipitous information. It is being used in interaction with testing, not as a substitute for it.

References

Headquarters, Department of the Army. (1991). Tactics, techniques, and procedures for forward arming and refueling points (Coordinating Draft of Field Manual 1-104). Washington, DC: Author.

McAnulty, D. M., Bierbaum, C., & Allender, L. (1992) Development of HARDMAN III models of forward arming and refueling point operations (ARI Research Report). Alexandria, VA: Army Research Institute.

MTA, Incorporated. (1989). Forward arming and refueling point (Final Report on Contract no. DAAA21-88-C-0112 Task 4). Huntsville, AL: Author.

Western Design Corporation. (1988). Modular ammunition package and feed system for attack helicopters (Final Report on Contract no. DAAA21-87-C-0275). Irvine, CA: Author.

ON MINIMIZING FRATRICIDE RISKS

Gilbert L. Neal

U. S. Army TRADOC Analysis Command
White Sands Missile Range, NM 88002

***Fratricide** is the employment of friendly weapons and munitions with intent to kill the enemy or destroy his equipment or facilities that results in unforeseen and unintentional death or injury to friendly personnel and/or damage or destruction of friendly materiel* (Modified from Dickson & Hundley, 1992 and Hennies & Dierberger, 1992). Friendly fire incidents also produce psychological aftereffects. Soldiers experiencing and surviving such events may experience reduced combat effectiveness that seriously affects unit ability to survive and function (Center for Army Lessons Learned [CALL], 1992a, 1992b). There are many reasons for friendly fire incidents (CALL, 1992a, 1992b). This paper, however, focuses only on preventing those due to target identification error.

Soldiers have feared friendly fire incidents and fratricide since the beginnings of organized warfare. Shrader (see Dickson and Hundley, 1992) estimated that friendly fire incidents accounted for two percent of the casualties in World Wars I and II, the Korean War, and the Vietnam War. During those and earlier wars such incidents have been well-documented. Highly publicized DESERT STORM friendly fire incidents renewed and accelerated interest in minimizing the risks of these tragedies, particularly those resulting from ground-to-ground and air-to-ground engagements. Some military analysts contend that advances in military technology are increasing rather than reducing fratricide risks (Dickson & Hundley, 1992). Proposed doctrinal, organizational, materiel, training, leadership, and risk management solutions to the problem have been described and discussed by CALL (1992a, 1992b), Dickson and Hundley (1992), and Hennies and Diersberger (1992). Materiel solutions, the point of departure for this paper, ranging from relatively simple visual identification systems to technically sophisticated identification friend or foe-type (IFF) systems, have been proposed to protect ground combat systems (CALL, 1992b; Dickson and Hundley, 1992).

Since World War II, electronic IFF systems have been used to protect friendly aircraft from friendly air defense artillery weapons. Generically, an IFF system is a secure coded electronic interrogate-response system, that uses frequently changed codes and whose use is governed by a set of strict operating procedures. For example, an operator/gunner challenges an approaching low flying aircraft using an encoded IFF signal. If the aircraft responds with the correct coded signal, the operator assumes it is friendly and does not engage it.

The results of post fielding training effectiveness analysis (PFTEA) studies (Neal, et al, 1991) of IFF training programs, strategies, and problem areas provide valuable insights and guidance applicable to insuring operational effectiveness of most battlefield identification systems proposed to protect infantry, armor, cavalry, field artillery, other ground combat, and combat support systems from friendly fire.

WORLD-WIDE JOINT IFF PFTEA

In 1989-90 we conducted a world-wide IFF PFTEA study (Vargas, Esparza, Howard & Zarret, 1990 and Vargas, Howard, Esparza, & Zarret, 1990) using a "closed loop approach" (Neal, et al, 1991) to assess the effectiveness of training programs and strategies for both forward area Army Air Defense Artillery (ADA) and Army Aviation (AVN) soldiers who perform Mark XII IFF system tasks. A basic premise of this study was that effective and successful operation of the IFF system is dependent on the task proficiencies of both ADA and AVN soldiers.

To insure that one Mark XII IFF system interrogation-response cycle is successfully completed and one friendly aircraft is protected, ADA and AVN soldiers, between them, must have performed 13 critical tasks -- six aircraft and seven ADA weapon -- error free. Associated with task performance are classified IFF codes that are typically changed every 24 hours; soldiers having to know and to understand Mode 3 (non-secure aircraft identification) and Mode 4 (cryptosecure identification of friendly aircraft) operations; and, rigorous operating procedures.

METHOD

Sample. Two soldier samples were assessed. (1) *ADA Sample.* This sample consisted of 825 ADA soldiers in four military occupational specialties (MOS) in skill levels 1 through 4 and areas of concentration (AOC) -- MOSs 16J, 16P, and 16S and AOC 14B. The soldiers were Stinger, Chaparral, and Forward Area Acquisition Radar crewmen and platoon leaders, battery officers and staff officers. (2) *AVN Sample.* This sample consisted of 365 AVN soldiers in eighteen MOSs, MOS series, and AOCs -- MOSs 35K, 35P, 35R, 93B, 93P; MOS series 67; MOSs 152B, C, D, F, & G; MOSs 153 A, B, C, & D; and, AOCs 15A, B, & C. Soldiers' jobs included: (a) enlisted repairers, operations coordinators, and aerial scouts; (b) warrant officer andD officer aviators; and, (c) staff officers. (3) *Common Factors.* The soldiers in both samples had IFF responsibilities and/or performed IFF tasks. Soldiers were assigned to ADA or AVN units stationed in the Continental United States and in U. S. Army Europe (USAREUR). Where feasible ADA and AVN units assigned to the same parent unit (e.g., a division) participated in the study.

Performance Assessment. Soldier IFF task proficiency and factors impacting proficiency were assessed as follows.

ADA Soldiers. (1) Soldiers in MOSs 16P and 16S who were designated IFF programmers and all MOS 16J soldiers each received a multiple choice multi-subject area IFF skills and knowledge (S&K) tests specific to their weapon system. (2) Each MOS also received weapon system specific "go no-go" scored IFF programming hands-on tests (HOT) consisting of seven scored subtasks (or eight steps) for MOS 16J and eleven scored subtasks (or 21 steps) for MOSs 16P and 16S. (3) All MOS 16Ps and 16Ss received an IFF Tone Recognition Test (Mode 3, Mode 4, or unknown) since the STINGER and CHAPARRAL IFF displays are auditory. (4) All soldiers were administered job, MOS, and AOC specific background and perception questionnaire sets that addressed training background and perceptions of IFF task criticality, command emphasis, use, training, proficiency, effectiveness, etc. (5) Officers received a paper-and-pencil interview questionnaire with questions concerning command emphasis, IFF system and training problems, and IFF system use.

AVN Soldiers. Assessment of AVN soldiers paralleled that of the ADA sample. (1) A seven subject area subtest S&K test was developed for Aviation IFF tasks. Each MOS/AOC was tested using an S&K test containing three common subtests plus subtests pertinent to MOS/AOC IFF responsibilities. (2) All MOSs and AOCs received a common "go-no go" scored IFF programming HOT. (3) Additional job specific IFF HOTs were administered to MOS 35K soldiers and to aviators. (4) All MOSs/AOCs received a soldier perception survey assessing IFF task criticality, frequency, and proficiency; factors impacting IFF critical task performance proficiency; availability, use, and ease of use of IFF manuals; and, institutional and unit training program effectiveness. (5) All soldiers received structured interviews addressing IFF operational use, unit procedures, and overall effectiveness. (6) A demographic survey assessed key soldier background and experience factors.

Joint Operational Test (OT). A limited controlled two-sided two-day exercise was conducted at one representative unit site to assess ADA crew and aircrew (helicopter) performance in using the complete IFF interrogation/response cycle. According to the test plan, each day, five STINGER teams, two CHAPARRAL crews, and one FAAR platoon were

positioned tactically on the ground. Five helicopters programmed in IFF Modes 3 and/or 4 flew scheduled scenario flight profiles over the ADA sites area. The ADA units were expected to interrogate each aircraft. After each flight, AVN data collectors (1) determined if aircraft transponders had been correctly programmed, and (2) debriefed flight crews to determine (a) if the aircraft had been interrogated and (b) if a valid IFF code had been transmitted. On the ADA unit side the data collectors assessed (1) correct programming of IFF interrogators, (2) correct interrogation of aircraft, and, (3) correct identification of IFF Mode 3, Mode 4, and unknown responses by soldiers.

RESULTS

S&K Tests. A S&K test score of 70 percent correct was a consensually established standard for minimum acceptable IFF task knowledge. Overall, the average S&K scores for the majority of the MOS/AOC samples were lower than the acceptable minimum.

ADA Soldiers. Mean S&K test scores for ADA soldiers were "low" to "minimum". Mean scores ranged from 54.6 (MOS 16J) to 70.1 (MOS 16S). MOS 16S soldiers were more likely to have subtest scores higher than 70 (i.e., 6 out of 8) than the other MOSs. In general, S&K scores increased with MOS skill level, but the differences were small and not statistically significant.

AVN Soldiers. Mean overall, common subject area, and subtest scores for AVN soldiers never exceeded the 70 percent minimum. Mean overall S&K scores ranged from 30.4 (MOS 67 series) to 52.4 (AOC 15). Officer (AOC 15) and warrant officer (MOS 152 & 153 series) aviators tended to have the higher S&K scores.

HOT Test (IFF Programming). *ADA Soldiers.* Specific ADA duty positions have IFF programming responsibilities. Eighty-six out of 99 MOS 16J soldiers attempted the "IFF programming" task; only 45.5 percent of all MOS 16Js performed all seven scored subtasks correctly. Of 82 MOS 16P soldiers, 72 attempted "IFF programming"; and only 4.8 percent completed all 11 scored subtasks correctly. Ninety-nine out of 114 MOS 16S soldiers attempted the task; only 1.8 percent could perform all eleven scored HOT subtasks correctly. IFF programming scores were also unrelated to MOS skill level.

AVN Soldiers. When the study was conducted, AVN IFF tasks could be assigned to any of the 18 AVN MOSs/AOCs in the study sample. For AVN soldiers the most important HOT task was "IFF Programming" (i.e., selecting the correct daily code from the code book, performing the correct IFF settings, and transferring the code to the computer). Overall, only 31.1 percent (N=245) of the AVN soldiers performed the programming task correctly. MOS 93P soldiers (N=22) performed least well (13.6 percent completed the task), and MOS 93B soldiers (N=29) performed best (49.0 percent completed the task). The percent aviators correctly completing the task were low -- MOS 152 (36.2 percent), MOS 153 (40.0 percent), and AOC 15 (42.0 percent).

HOT (Tone Recognition Test). The CHAPARRAL and STINGER systems have aural IFF displays. As part of the target engagement decision process MOS 16P and MOS 16S gunners must recognize three different IFF response tones -- Mode 3, Mode 4, and Unknown. MOS 16P (N=218) percent correct tone recognitions were: Mode 3 (52.3); Mode 4 (52.5); and Unknown (77.8). MOS 16S (N=343) percent correct tone recognitions were: Mode 3 (79.1); Mode 4 (80.7); and, Unknown (92.3).

Joint Operational Test. (1) *Day 1.* (a) Only one STINGER team of five was correctly coded for Modes 3 and 4. (b) The one team, in addition to correct IFF responses, experienced Mode 3 and Unknown responses due to pointing errors and antenna masking effects. (c) Four aircraft out of five had correct IFF codes and recorded Mode 4 IFF interrogations. (d) Equipment malfunctions precluded data from the FAAR and CHAPARRAL teams. (2) *Day 2.* (a) Only one out of four

STINGERS was correctly coded for Mode 4. (b) Only one out of four aircraft had correct Modes 3 and 4 codes. (c) Equipment problems again limited expected data.

Perception and Interview Data. *ADA Soldiers.* (1) (a) Over 80 percent of each of MOS 16J, 16P, and 16S soldiers rated their own IFF programming abilities as "good". (This was not supported by test scores!) (b) MOS 16S soldiers more accurately assessed their own IFF tone recognition abilities than did MOS 16P soldiers. (c) A majority (over 80 percent) of all MOSs rated IFF tasks "critical" and believed that use of IFF would decrease fratricide possibilities. (d) Slightly more than a third of the soldiers perceived that their commanders placed command emphasis on proper IFF use. (e) Overall, soldiers reported confidence in their IFF equipment and would use it in combat (73 percent), because it would reduce fratricide (65 percent). (f) However, a third were concerned that use of IFF would reveal their positions to the enemy. (g) All MOSs reported concerns about IFF equipment shortages, school communications security training they received, reliability of equipment, and lack of IFF equipment in units. (h) Although exportable training materials, training films and special texts existed, less than nine percent of the 16P and 16S programmers reported having seen the film and almost as few programmers had instructors who used the special text. (2) Officers were more positive, but had major training concerns. (a) They were more likely to report: IFF equipment working properly (66.7 percent); fewer concerns about equipment availability (23 percent) and equipment capabilities and limitations (41 percent); confidence that the IFF subsystem would support the Air defense mission (68.2 percent) and would reduce fratricide (80.3 percent); their unit had adequate plans and procedures to support IFF (71.2 percent); and, that plans included joint training with Army Aviation (48 percent). (b) However, 36.4 percent reported they never trained with Army Aviation and used Mode 4, but when such training did occur, it could be monthly to annually. (c) Furthermore, more than 70 percent reported that IFF interrogators were not programmed during practice alerts, and a quarter to a half of their operators never programmed their interrogators nor interrogated an actual aircraft. (d) Lack of IFF practice with the Air Force and Army Aviation was a major concern (64 percent) and that fact adversely impacted training (55 percent).

AVN Soldiers. (1) Overall, 67.2 percent of the soldiers (N=365) reported that they believed their units had plans and procedures to employ IFF on a regular basis. (2) They also reported that joint AVN and ADA training was not conducted as a matter of policy (72.2 percent), and unit training programs were not adequate to sustain IFF system operational proficiency (70.1 percent). (3) Most soldiers believed that unit IFF operating procedures require training. (4) In general, soldiers failed to understand the basic IFF system fundamentals. Soldiers believed: (a) the Mode 4 IFF system had a higher security classification than it actually has (83 percent). (This discourages training use.); (b) IFF identifies hostile aircraft (54 percent). (It identifies "friendlies".); (c) IFF will respond with the wrong code inserted; and, (d) Mode 4 IFF enhances aircraft survivability (85 percent). (5) Overall, AVN soldiers had less than expected confidence (61 percent) in Army airspace command and control, but nearly one half (45 percent) believe that training in the area would reduce fratricide. (6) The majority of soldiers (except MOSs 67 and 93P) perceived IFF tasks to be mission critical, and stated they performed such tasks at least once quarterly.

Observation Data and Follow-up Investigations. Conclusions drawn from observations made by study team members and study excursions follow. (1) Perceived and actual restrictions on Mode 4 use in USAREUR discouraged unit training use, field training exercises use, and restricted data collection. (2) Soldiers did not fully understand and had misconceptions about IFF tasks and equipment. (3) Soldiers tended to overclassify IFF system components; such perceived classification levels discouraged unit training use. (4) Formal IFF individual and collective training was not, in fact, conducted by units. (5) When ADA and AVN units did practice together, they had neither standing operating procedures (SOP) to follow nor standardized methods to evaluate individual or joint exercise performance. (6) Both instructors and students had low IFF skills and knowledge proficiencies. (7) Mode 4 sustainment training was not

available for new unit assignees. (8) Maintenance problems reduced IFF equipment availability for school training.

CONCLUSIONS AND IMPLICATIONS

The PFTEA findings indicated that (1) soldier IFF task proficiency was low, and (2) suggest that a high risk of fratricide existed when this study was conducted even though a sophisticated fratricide prevention system was in use. The study provided the data needed to correct the problems. The results showed that a "high tech" fratricide prevention system will not perform as intended unless soldiers understand it and know how to operate it. Study findings reinforce the roles of training, policies, and total system design in the effective employment of any battlefield identification system. The findings support the need to apply the following principles to insure effective employment of such systems. (1) Soldiers and units must train to use these systems. (2) Such training should be frequent. (3) A standardized performance evaluation system must support this training. (4) Two-sided ("closed loop") training is essential to provide performance effectiveness feedback. (5) Command emphasis must be placed on such training. (6) Command policies must not unduly restrict this such training. (7) Security considerations should not discourage training use. (8) If there are security considerations, unclassified surrogate or simulator systems should be used in training. (9) Systems, procedures, job aids, and support equipment must be as easy to use a possible ("user friendly") to preclude operator error. (10) Critical task responsibilities should be clearly assigned to specific MOSs or duty positions.

REFERENCES

Center for Army Lessons Learned (1992a, March). *Fratricide: risk assessment for company leadership* (Handbook 92-3). Fort Leavenworth, KS: Combined Arms Command.

Center for Army Lessons Learned (1992b, April). *Fratricide: reducing self-inflicted losses* (Newsletter 92-4). Fort Leavenworth, KS: Combined Arms Command.

Dickson, D. D. Jr., & Hundley, E. (1992, July). Avoiding not so friendly fire. *Military Review. LXXII*(7), 57-64.

Hennies, C. A. & Dierberger, P. A. (1992, May). Risk management, a key battlefield edge. *Military Review. LXXII*(5), 32-40.

Neal, G. L., Catherson, N. S., Esparza, H. J., Howard, F. S., Zamarripa, A., & Vargas, A. A. (1991). Post fielding training effectiveness analysis: in remembrance. In *Proceedings, 33rd Annual Conference of Military Testing Association, San Antonio, TX, 28-31 October 1991* (pp. 722-726). San Antonio, TX: Armstrong Laboratory Human Resources Directorate/USAF Occupational Measurements Squadron.

Vargas, A. A., Howard, F. S., Esparza, H. J., & Zarret, J. (1990, October). *Joint Air Defense Artillery/Army Aviation Identification Friend or Foe Post Fielding Training Effectiveness Analysis, Volume II -- ADA IFF PFTEA* (TRAC-WSMR- PTEA-91007-02). White Sands Missile Range, NM: U. S. Army TRADOC Analysis Command-WSMR.

Vargas, A.A., Howard, F.S., Esparza, H.J., & Zarret, J. (1990, October). *Joint Air Defense Artillery/Army Aviation Identification Friend or Foe Post Fielding Training Effectiveness Analysis, Volume III -- AVN IFF PFTEA* (TRAC-WSMR-PFTEA-91-007-03). White Sands Missile Range, NM: U.S. Army TRADOC Analysis Command-WSMR.

DISCLAIMER: The views expressed in this paper represent those of the author and do not necessarily represent the views, position, or policy of the U. S. Army.

ACKNOWLEDGEMENTS: The author gratefully acknowledges the constructive critiques by Sylvia C. Acchione-Noel, Fay S. Howard, Charles R. Hughes, and Alfonso A. Zamarripa and the technical support provided by Soyla V. Barreras and Patricia Fulton.

OCCUPATIONAL SURVEY REPORT

Education and Training Utilization Field, Special Duties, and Technical Instructors

Capt Harold Huguley III, USAF Occupational Measurement Squadron

ABSTRACT

This report presents the results of a detailed United States Air Force Occupational Survey of the Education and Training Utilization Field (AFSC 75XX), four Special Duty Identifiers (SDI) (0900, 0940, 0950, and 0970), and Technical Instructors (T-Prefix officers). This study objectively identifies and describes the AFSC 75XX and SDI 09XX jobs, and describes training management and development functions across all AFSCs using the T-Prefix personnel. The sample population consists of 2,222 officers, which represents 62 percent of the eligible population. There were 79 Education and Training jobs identified. Sixty-eight of the jobs grouped together to form 11 clusters, and the other jobs were identified as independent jobs. In addition, comprehensive analysis was conducted on first-assignment personnel, military rank, duty AFSC, and job satisfaction. This report was used to assist in personnel and training management decisions.

The AFSC 75XX utilization field originated in 1954 as three AFSs; Education and Training Staff Officers (AFS 751X), Education Specialist (AFS 752X), and Instructors (AFS 753X). In 1960, AFS 752X was renamed Education and Training Officers. In 1970, AFS 753X became SDI 0904 - Instructors; SDI 0904 was redesignated SDI 0940 in 1974.

As described by AFR 36-1 Officer Classification, the current Education and Training Utilization Field (AFSC 75XX) encompasses the functions and responsibilities of planning, organizing, establishing, and directing education and training programs. The entry-level officer AFSC is 7521/7524, and the rank spread stated in AFR 36-1 is second lieutenant through major. Staff officer AFSC is 7511/7516, and the rank spread stated in AFR 36-1 is major through colonel.

SDI authorizations for officers identify personnel who are performing an actual group of tasks on a semipermanent or permanent basis. These duties are unrelated to any specific utilization field. The officers who possess the SDI 0900 are commanders for the USAF Academy Cadet Squadrons. They are responsible for commanding the cadet squadrons; directing actions appropriate to morale, welfare, discipline, and aptitude; and coordinating training and instruction programs. AFR 36-1 states the rank spread for SDI 0900 is captain and major.

The officers who possess the SDI 0940 identifier are formal instructors. They are responsible for organizing and preparing instructional materials, instructing personnel, and coordinating training programs. AFR 36-1 states the rank spread for 0940 SDIs is second lieutenant through colonel.

The 0950 SDI are Training Commanders at Officer Training School (OTS). Their responsibilities include commanding training squadrons and flights, determining aptitude for commissioned service, and directing and maintaining training programs. AFR 36-1 states the rank spread for SDI 0950 is first lieutenant through major.

The 0970 SDI are Academic Program Managers. They are responsible for directing, instructing, evaluating, and monitoring all instructions, curriculum development, and student training at the USAF Academy and Professional Military Education (PME) schools. AFR 36-1 states the rank spread for SDI 0970 officers is captain through colonel.

The T-Prefix identifies officers serving in positions as instructors in technical subjects. It applies to nonrated specialties and in-ground phases of pilot and navigation specialties. It is awarded and affixed to the awarded AFSC in which the officer performs duty as a technical instructor.

The data collection instrument for this occupational survey was "Education and Training Officer Personnel USAF Job Inventory, AFPT 90-75X-911," dated March 1990. The inventory consisted of two main sections; the respondents' biographical and current job information section and a detailed list of tasks performed at all organizational levels of education and training, special duties, and technical instruction. The task list was tentatively prepared after reviewing the two previous job inventories, the education and training publications, and all pertinent directives. The list was further developed by selected Subject-Matter Experts (SME) at Keesler AFB, Maxwell AFB, Randolph AFB, and Lackland AFB.

During July through November 1990, 3,549 Education and Training Job Inventories (JI) were administered in an effort to capture all eligible education and training personnel. In total, 2,222 JIs were returned and analyzed; this represents 62 percent of the 1990 eligible population.

REPRESENTATION OF 75XX, 09XX AND T-PREFIX
PERSONNEL WITHIN SURVEY SAMPLE

AFS	ELIGIBLE POP	% of TOTAL ELIGIBLE POP	NUMBER RETURNED	% OF TOTAL RETURNED	% OF ELIGIBLE AFS	% OF TOTAL ELIGIBLE POP
75XX	414	12%	259	12%	63%	7%
09XX	1,388	39%	910	41%	66%	25%
T-Prefix	1,747	49%	1,053	47%	60%	30%
Total	3,549	100%	2,222	100%		62%

All individuals who filled out an inventory completed an identification and biographical section. Next, they went through the booklet and checked each task performed in their current job. Finally, they went back and rated each task they checked on a 9-point scale reflecting relative time spent on each task compared to all other tasks. Ratings range from "1," which indicated very small amount of time spent, to "9," which indicated a very large amount of time spent. The relative percent time spent on tasks for each inventory was captured by first totaling all rating values on the inventory. The rating for each task was then divided by this total and the result multiplied by 100. The percent time spent ratings from all inventories were combined and used with percent members performing to describe the various groups in the career utilization field.

TE booklets were completed by 141 experienced AFSC 75XX, SDI 09XX, or T-Prefix officers in the grades of captain through lieutenant colonel. Individuals completing the TE booklets were also asked to rate tasks on a 9-point scale (from "1," no training is required, to "9," extremely high amount of training needed). The TE rating is a relative comparison of which tasks require structured training of new education and training personnel (first 48 months in the career field). "Structured" training is defined as training provided at training schools, field training detachments, mobile training teams, formal on-the-job training, or any other organized training method. For this survey, the responses indicate an average (mean) TE of .76; however, some tasks are as high as 5.319.

Once the job inventories and the task factor booklets were received from the field, a very powerful computer, written to analyze occupational input data called the Comprehensive Occupational Data Analysis Program (CODAP), created a job description for each respondent, as well as composite job descriptions for members of various demographic groups.

For the purpose of organizing individual jobs into similar units of work, CODAP used an automated job clustering process. The basic identified group in this hierarchial process is referred to as a "Job." If this job has distinguishing characteristics which are unrelated to other jobs, it is referred to as an "Independent Job." When there is a substantial degree of similarity between jobs, they are grouped together and identified as a "Job Cluster." The resulting data may be used to evaluate the accuracy of career documents (e.g., AFR 36-1) and to gain a better understanding of current personnel utilization and training applications.

There are 79 primary education and training jobs which are identified; this represents 80 percent of the education and training survey. The majority of the respondents spend their duty time performing tasks associated with General Command, Staff, and Administrative Functions (22 percent); Conducting Education or Training (19 percent); and Developing Courses, Curriculum, and Course Documents (18 percent). Consequently, there is a high overlap (a set of common core tasks) among most education and training jobs.

An examination of DAFSC groups, along with the analysis of identified jobs, is an important part of each occupational analysis. The DAFSC analysis reveals similarities and differences among various levels, based on tasks they performed and the relative time spent on particular duties. The information is used to assess the accuracy and how well the utilization field document (AFR 36-1 Specialty Descriptions) reflect what career ladder personnel are actually doing in the field.

As the officer entry-level DAFSC 7521, the personnel in this group devote the majority of their duty time to performing administrative functions. The average number of tasks performed by DAFSC 7521 officers is 26. The 24 officers in this group spend 67 percent of the time Performing General Command, Staff, and Administrative Functions; 9 percent of their time is spent Developing Courses, Curriculum, and Course Documents; 8 percent of their time is spent Evaluating or Inspecting Education or Training Programs; and 8 percent of their time is spent Performing Supervisory and Personnel Staff Functions. The majority (39 percent) of the DAFSC 7521 officers were in staff jobs. Some of the officers enter the utilization field after serving in a previous DAFSC; this is reflected by the DAFSC 7521 rank distribution, which includes second lieutenants (40 percent), first lieutenants (4 percent), captains (30 percent), majors (17 percent) and lieutenant colonels (9 percent). The duty descriptions written in AFR 36-1 is accurate; however, it is more common for lieutenant colonels to occupy DAFSC 7516.

The upgraded DAFSC 7524 performs more tasks commensurate with their gained knowledge and rank. The 90 officers in this group spend 72 percent of their duty time performing General Command, Staff, and Administrative Functions; they spend 15 percent of thier duty time Developing Courses, Curriculum, and Course Documents; they spend 10 percent of their duty time Performing Supervisory and Personnel Staff Functions; and 7 percent of their duty time Determining Education and Training Requirements. The majority (32 percent) of the officers work in Command and Staff jobs, while 13 percent are Curriculum Developers, and 10 percent are Plans and Programs Officers. The rank includes first lieutenants (18 percent), captains (72 percent), majors (8 percent), and lieutenant colonels (2 percent). This indicates a majority of the DAFSC 7521 members are upgrading into the DAFSC 7524 DAFSC; however, after the rank of captain is achieved, they either separate or transition to DAFSC 7511 or cross-train into some other DAFSC. The duty description written in AFR 36-1 is accurate; however, it is more common for lieutenant colonels to occupy DAFSC 7516.

The analysis indicates there is evidence of a slight career progression pattern for DAFSC 7521/7524 officers who advance to become DAFSC 7511 officers. The 27 members of this group perform an average of 61 tasks. They spend 70 percent of their duty time Performing General Command, Staff, and Administrative Functions; they spend 13 percent of their duty time Performing Supervisory and Personnel Staff Functions; and they spend 9 percent of their duty time Developing Courses, Curriculum, and Course Documents. Nearly two-thirds of the members (65 percent) work in Command and Staff jobs, while 17 percent work as Plans and Programs Officers, and 9 percent are Faculty Administrators. The rank spread of AFSC 7511 officers includes captains (37 percent), majors (30 percent), lieutenant colonels (22 percent), and colonels (11 percent). The description written in AFR 36-1 is accurate; however, it is more common for captains to occupy DAFSC 7524.

The DAFSC 7516 is the largest and most senior-ranking in the AFSC 75XX. The 119 members perform an average of 80 tasks. They spend 63 percent of their time Performing General Command, Staff, and Administrative Functions; they spend 15 percent of their duty time Performing Supervisory and Personnel Staff Functions; and they spend 11 percent of their duty time Developing Courses, Curriculum, and Course Documents. More than half (58 percent) hold Command and Staff jobs, while 13 percent are Faculty Administrators. The rank spread included captains (7 percent), majors (37 percent), lieutenant colonels (36 percent) and colonels (20 percent). The description written in AFR 36-1 is accurate; however, it is more common for captains to occupy DAFSC 7524.

The SDI 0900 was a relatively small group with 13 members. They perform an average of 145 tasks. These members spend 76 percent of their duty time Managing or Counseling Students; they spend 24 percent of their duty time Conducting Education or Training; and they spend 23 percent of their duty time Performing General Command, Staff, and Administrative Functions. This is a specific job held by officers who command cadet squadrons at the USAF Academy. The rank spread includes captains (46 percent) and majors (54 percent). The duty descriptions written in AFR 36-1 is accurate.

The SDI 0940 was the largest SDI group with 710 members. They perform a average of 129 tasks. These members spend 71 percent of their duty time Conducting Education or Training; they spend 20 percent of their time Performing General Command Staff and Administrative Functions; and they spend 12 percent of their duty time Developing Courses, Curriculum, and Course Documents. Approximately half the members (47 percent) hold Management and Counseling jobs, while 16 percent have Faculty Instructor jobs, and 9 percent are Training Evaluators. The rank spread includes second lieutenants (1 percent), captains (66 percent), majors (17 percent), lieutenant colonels (7 percent), and colonels (9 percent). The duty descriptions written in AFR 36-1 is accurate.

There were 49 members in the SDI 0950 group. They performed an average of 79 tasks. They spend 67 percent of their duty time Conducting Education or Training; they spend 24 percent of their duty time Performing General Command, Staff, and Administrative Functions (24 percent); and

they spend 14 percent of their duty time Managing and Counseling Students. Most members (29 percent) are Training Evaluators, while 16 percent hold Command and Staff jobs, and 10 percent are Military Training School Commanders. The rank spread includes captains (90 percent) and majors (10 percent). The description written in AFR 36-1 is accurate.

The SDI 0970 was the second largest SDI group with 136 members. They performed an average of 117 tasks. These members spend 79 percent of their duty time Performing Command, Staff, and Administrative Functions; they spend 18 percent of their duty time Developing Courses, Curriculum, and Course Documents; and they spend 15 percent of their duty time Conducting Education or Training. The jobs most likely associated with SDI 0970 are Faculty Instructors (17 percent), Faculty Administrators (16 percent), and Command and Staff (12 percent). The rank spread includes second lieutenants (2 percent), captains (30 percent), majors (33 percent), lieutenant colonels (24 percent), and colonels (11 percent). The duty descriptions written in AFR 36-1 is accurate; however, second lieutenants should not have been considered for this special duty.

There were 1,055 members identified who held a T-Prefix. They performed an average of 110 tasks. The members spend 62 percent of duty time Developing Courses, Curriculum, and Course Documents; they spend 19 percent of their duty time Performing Command, Staff, and Administrative Functions; and they spend 18 percent of their duty time Conducting Education and Training. The most common jobs for T-Prefix officers include Faculty Instructors (24 percent), Command and Staff (15 percent), Education and Training Instructors (11 percent), and Faculty Administration (11 percent). The rank spread included second lieutenants (1 percent), first lieutenants (4 percent), captains (60 percent), majors (20 percent), lieutenant colonels (11 percent), and colonels (4 percent).

In performing an analysis of the military rank, lieutenants were generally performing a high degree of administrative and staff duties. Although captains also perform administrative tasks, they spend a lot of time conducting or developing education and training programs. Survey respondents who were of the rank major through colonel perform duties associated with command and staff. They occupy leadership positions within the education and training utilization field.

The survey included 1,339 respondents who were first-assignment personnel. First-assignment 75XX personnel mostly performed administrative and staffing duties; first-assignment 09XX officers are usually conducting education and training; and first-assignment T-Prefix officers generally develop courses and curriculum.

Training Emphasis (TE) ratings are factors that can assist technical school personnel in deciding which tasks should be emphasized for entry-level training. In addition, it may provide support for adding or deleting training requiements. The TE ratings provided by the education and training SMEs yielded an average (mean) rating of .76 with a standard deviation of 1.76. According to ATCR 52-22, when a given task has an assigned TE rating greater than or equal to the sum of the mean value plus one standard deviation, in this case 2.52, it merits strong consideration for inclusion in some for of structured training. Only 41 of the 968 tasks met this criteria.

The job satisfaction indicators show moderate to high levels of satisfaction among education and training jobs, as well as across DAFSCs. There is indication that talents and training are being undertutilized; this is evident across clusters and independent jobs and DAFSCs.

Overall, officers involved with education and training were fairly satisfied. The two lowest percentages for job interest were AWC Curriculum Developers (67 percent) and Plans and Programs Officers (57 percent). Fundamental Training Instructors and Plans and Programs Officers perceive the use of their talents to be low. More than half of the Plans and Programs Officers (51 percent) also feel use of training is low. All personnel working as Military Training School Commanders, Research Directors, Foreign Military Training Officers, and AFIT Research Professors are completely satisfied with their work accomplishment. Two independent jobs gave indication of opportunities for overseas assignments; they are Liaison/Public Affairs Officers with 67 percent and Foreign Military Training Officers with 60 percent. Management and Counseling Officers spend twice as much or more time performing additional duties than any other cluster or independent job. As for career plans and intentions, most of the personnel involved with education and training prefer to stay in the Air Force and retire. However, some officers do show interest in cross-training or changing their AFSC.

The job satisfaction across DAFSC is also moderate. AFSC 75XX, SDI 09XX, and T-Prefix personnel, as a whole, express interest in their job and are satisfied with their work accomplishment; however, all three groups perceive their talents and training to be underutilized. There are very few overseas opportunities in any of the groups. The SDI 09XX personnel tend to spend twice as much time on additional duties than do AFSC 75XX or T-Prefix personnel. The career intentions for all three groups is about the same with most members, indicating a desire

to stay and retire with benefits. Career plans show that roughly 60 percent of the AFSC 75XX and T-Prefix personnel want to stay in their current AFSC; only 34 percent of the SDI 09XXs want to stay in their current AFSC. This should be expected since the SDI 09XX utilization field is a special duty assignment.

Personnel involved with education and training are very diverse. No one single job can accurately depict the field. They perform many functions and duties spanning 89 jobs; the average number of tasks ranged from 20 to 240 per job; and their time spent fluctuates across most duties. In all jobs occupied by AFSC 75XX officers, such as Faculty Instructors, Faculty Administrators, and Command and Staff, SDI 09XX, and/or T-Prefix personnel, were found performing the same duties and tasks. Although the AFSC 75XX, SDI 09XX, and T-Prefix personnel often perform the same job, AFSC 75XX officers are assigned education and training jobs as career jobs, while SDI 09XX and T-Prefix personnel are assigned education and training jobs for career broadening. This often causes disillusionment to career-oriented AFSC 75XX personnel. A new clear, comprehensive career plan, outlining the future needs and purposes for all education and training personnel, ought to be developed.

AFR 36-1 Specialty Descriptions were generally consistent with the actual duties and tasks performed by DAFSC 75XX and SDI 09XX personnel across all jobs. In addition, the duties performed by T-Prefix personnel correlated with typical duties which would be expected to be performed by technical instructors. However, the majority of the AFSC 75XX officers perform tasks associated with staff and administrative functions; these tasks can generally be accomplished by any Air Force officer. This is especially true of DAFSC 7521 officers who are primarily performing administrative intensive tasks. Other AFSC 75XX officers are in jobs which have duties associated with curriculum development, supervisory, and personnel functions; these duties would definitely require adequate experience and training. However, these duties and the percent time spent performing the duties still remain a small facet of the overall career field.

Out of the 79 identified jobs, only 4 were considered as senior leadership jobs. Two were independent jobs, Research Directors and AWC Curriculum Developers. The other two were included in job clusters, USAFA Department Heads/Deputies (Faculty Administration), and Unit Commanders/ Directors (Command and Staff). Lieutenant colonel was the predominant grade held by officers in these jobs. Although 20 percent of AFSC 75XX officers held leadership jobs, the number of leadership jobs given to SDI 09XX and T-Prefix officers was actually three to four times greater than the number given AFSC 75XX officers.

Many DAFSC 7521 officers are selected for jobs unrelated to their training and expertise. This contributes to their low job satisfaction. As the DAFSC 7521 officers upgrade to DAFSC 7524 and the 751X specialties, their job satisfaction increases 20-30 percent--most likely due to performing more duties directly related to education and training. Although advanced AFSC 75XX officers perform more education and training duties, there are still many AFSC 75XX officers performing duties unrelated to education and training. In addition, many AFSC 75XX officers feel their training and talents are underutilized. SDI 09XX and T-Prefix personnel seem generally satisfied with the jobs they are performing. Although they also indicated their training and talents were underutilized in their current job, this would be expected for officers in a special duty or career-broadening assignment. It is imperative to not only identify true education and training jobs, but also to match these jobs with the appropriately qualified personnel.

Although the jobs are numerous and vary according to duties and tasks, officers would benefit greatly from an overall strategic career plan which specifically outlines their role. In establishing a new career plan, AFR 36-1 would suffice as a guide for duties and responsibilities. A new career structure might facilitate an expansion of leadership opportunities for AFSC 75XX officers, as well as provide an accurate means of matching qualified officers with education and training jobs.

UTILITY OF OCCUPATIONAL SURVEYS FOR ASSESSING ENLISTED JOB PERFORMANCE

1Lt Mark R. Miller
Jacobina Skinner
Donald L. Harville

Armstrong Laboratory Human Resources Directorate
Brooks Air Force Base, Texas

Job performance is frequently measured by supervisory ratings, though the ratings are often criticized for such problems as unreliability, leniency, range restriction, and rater-ratee favoritism. During the 1980s, the Air Force worked to develop improved job performance measures (Hedge & Teachout, 1986). Several techniques including job knowledge tests, walk through performance tests, interview tests, and a variety of rating forms were attempted. Promising levels of reliability and validity were achieved. However, measurement costs proved to be a major drawback; test development and data collection were labor intensive and time consuming. The Air Force's contribution to the Joint Services Job Performance Measurement (JPM) Project cost about \$.5 to \$1 million for each of the eight specialties studied.

In the present investigation, preexisting Air Force occupational surveys were examined as a low cost, alternative source for job performance indices on enlisted personnel. The design for the study drew on a conceptualization of job performance proposed by Nathan (1992). He suggested that given the opportunity, managers will assign tasks to those employees most capable of performing them. By inference, the scope and complexity of tasks performed should increase for higher aptitude and for more experienced employees. The current study explored whether the expected patterns were revealed in indices based on occupational surveys of enlisted specialties.

Nathan (1992) explained his concepts by addressing supervisory ratings as a validation measure. He challenged the assumption made by researchers using supervisory ratings that job incumbents in a given type of job perform the same tasks, but with differing competence as a function of each incumbent's ability. If Nathan is correct, the positions that make up the same job would not be identical. Differential task assignment patterns would confound supervisory ratings of performance and limit their utility as criteria for validating personnel selection tests. With these issues in mind, Nathan (1992) investigated the number of tasks performed by 130 clerical workers. The validity coefficients for aptitude tests were consistently higher with task counts than with ratings of job performance, and Nathan's hypothesis was confirmed.

The present investigation expanded on Nathan's evaluation of job characteristics as job performance criteria. In addition to number of tasks performed, aptitude relationships with two measures of job difficulty, available from Air Force occupational surveys, were examined. Further, job experience, a potential contaminant in Nathan's (1992) study, was controlled. Prior Air Force research shows that task counts and difficulty increase for incumbents over time (Christal, 1974). A further distinction concerned study goals. Nathan focused on the usefulness of job characteristics as validation criteria for personnel selection tests. In contrast, the present investigation was spurred by Air Force needs to examine aptitude and experience effects in the context of work force structuring. Previous research using job performance measures (Alley & Teachout, 1990) was limited to the eight specialties in the JPM project. If occupational survey data can be exploited for criteria, work force structure issues could be explored on a broader scale. Costs would be minimal, because large numbers (about 50) of Air Force enlisted jobs are surveyed annually.

METHOD

Choice of Occupational Area. Twenty-three (23) specialties surveyed between 1985 and 1991 were selected for the study (see Table 1). Factors considered in choosing specialties for analysis were: 1) recency of the occupational survey, 2) flightline operations mission rather than administrative/support mission, and 3) relatively large numbers of personnel surveyed to maximize confidence in the findings.

Subjects. Subjects were 13,117 Air Force enlisted personnel who completed the 23 job inventories. In order to concentrate on the technical tasks, superintendents (9-skill level), who perform mainly administrative and management tasks, were omitted. The number of subjects in each specialty ranged from 183 to 2,011, with a mean of 570 (see Table 1).

Task Oriented Performance Measures (TOPMs). Three performance measures were computed from job incumbents' responses to the inventories: 1) Number of Tasks Performed (NTP); 2) Average Task Difficulty Per Unit Time Spent (ATDPUTS); and 3) Job Difficulty Index (JDI). Most job inventories contained 500 task statements or more, and were designed to describe completely the work performed by entry-level apprentices through senior enlisted managers. The job incumbent marked those tasks which he/she actually performed, and the total count was taken as the NTP criterion. The ATDPUTS criterion was obtained from two task factors, a "time spent" factor and a "task difficulty" factor. Incumbents rated how much work time they devoted to each task actually performed on a 9-point "relative time spent scale" (from very small to very large). For each incumbent the ratings were converted to indicate the percentage of work time spent on each task. Samples of supervisors rated the difficulty of each task on a 9-point scale describing the amount of time required to learn each task relative to other tasks in the inventory (from extremely low to extremely high). Their ratings for each task were averaged, and then standardized across tasks to a mean of 5.0 and a standard deviation of 1.0. An ATDPUTS value for each incumbent was computed by summing the cross-products of the "time spent" and "task difficulty" factors. The JDI criterion was a composite measure, derived from a regression analysis capturing supervisors' policy regarding the difficulty of jobs (Koym, 1977). Their judgments revealed that more difficult jobs have more tasks and higher difficulty tasks. The job difficulty prediction equation developed and validated by Koym for cross-specialty use read: expected JDI = (1.4431)(NTP) + (-0.8236)(NTP^2) + (0.4010)(ATDPUTS) for standard score (z-score) predictors. The correlation with supervisor ratings was above .81, a value which compared closely to results for specialty-specific equations. Koym's equation was used in the present study to obtain a predicted JDI for each subject.

Predictors. Prior to enlistment in the Air Force, each subject completed the Armed Services Vocational Aptitude Battery (ASVAB) as part of the screening process for military service. Minimum qualification scores, referred to as an Aptitude Index (AI), are established for Air Force jobs on four ASVAB classification composites: Mechanical (M), Administrative (A), General (G), and Electronic (E). Percentile scores achieved by subjects on the requisite classification composite for each job were obtained from historical files maintained by the Air Force Armstrong Laboratory. A second predictor, number of months of job experience, was obtained from job inventory background information provided by subjects.

Analysis. To evaluate sample performance, descriptive statistics (mean, standard deviation) were computed for criteria and predictor variables. Simple correlations were used to explore relationships among incumbent aptitude, job experience, and the amount and complexity of work performed, as reflected by the TOPMs. Correlations were not corrected for range restriction due to prior selection on the ASVAB. In addition, multiple regression analyses (linear models) were used to examine the joint effect of aptitude and experience. A full model containing both predictors was compared to a restricted

model containing the experience predictor alone. The difference in model R^2's (squared multiple correlation coefficients) was tested using the F statistic to determine whether aptitude made a unique contribution to TOPM prediction, when experience was held constant. For specialties with significant aptitude effects, over and above experience, expected criteria values (predicted scores) were inspected to determine performance gains and tradeoffs.

RESULTS

The samples were found to be highly selected on the aptitude variables, relative to American youth eligible for military service. Mean aptitude for the samples ranged from 67 to 92, about 30 to 40 percentile points higher than the average score (50.0) for the ASVAB reference group. Further, subjects varied considerably in terms of job experience, with means ranging from about 25 to 50 months and standard deviations from about 18 to 34 months across specialties. Descriptive statistics for the criteria also suggested substantial variability in most specialties. The ranges of means (and of standard deviations in parenthesis) for ATDPUTS were 4.2 to 5.4 (.2 to .4), and for NTP were 50.2 to 377.8 tasks (29.9 to 218.4). The predicted JDIs, which were measured in z-scores, had means equal to 0.00 and standard deviations which ranged from .8 to 1.1 across specialties.

Experience effects were detected consistently across specialties for the three performance criteria (see bivariate correlations in Table 1). Significant relationships were detected for at least one TOPM in all 23 specialties and for all three TOPMs in 17 specialties (74%). The correlation coefficients were mostly positive and moderate in size, typically in the .20 to .40 range. The few negative correlations in Table 1 were usually for specialties with smaller sample sizes. Overall, results indicated that the nature of the incumbents' work tended to change with the length of time assigned to the specialty, with variety (NTP) and difficulty (ATDPUTS and JDI) of work performed usually increasing for more experienced personnel.

The bivariate correlations for the ASVAB variables showed that effects due to aptitude alone were less salient than those observed for the job experience predictor (see Table 1). This result may be largely due to restriction in range from explicit selection effects. Significant positive relationships were found for at least one TOPM in over half the specialties (14 of 23). For three specialties aptitude was correlated with all three TOPMs.

Results of F-tests for an aptitude effect, with experience treated as a covariate, are shown in Table 2. Multiple correlations (R) for the full model and its R^2 difference, when compared to the restricted model, are reported. Unique aptitude effects (R^2 differences significant at $p < .05$ or lower) were detected for the 14 specialty/criterion combinations tabled. Most of the joint relationships (12 of 14) were in the positive direction, as illustrated in Figure 1, with performance increasing as experience and aptitude increased. The magnitudes of the performance gains were evaluated by computing expected criteria values (predicted scores) for each specialty, and then calculating the mean number of years of job experience and of aptitude percentile points needed to achieve a .5 standard deviation unit increase in the criteria. For job experience, the means were 4.6, 4.4, and 5.3 years for ATDPUTS, NTP, and JDI, respectively. The corresponding means for aptitude were 60, 55, and 63 percentile points. About 16 to 22 aptitude points, on average across specialties within criteria, were equivalent to 1 year of job experience.

DISCUSSION

The finding of simple aptitude effects for almost half of the specialties examined supported Nathan's

(1992) idea that job performance is not just how well one performs, but what one performs. In selected specialties, higher ability airmen performed work of slightly greater scope and/or complexity than that performed by lower ability airmen. The magnitudes of the bivariate correlations obtained were, however, smaller than those reported by Nathan for the clerical job, which ranged from .15 to .53. Reasons are unclear, but may include larger restriction in range, more heterogenous jobs within the same specialty, and curvilinear, rather than linear, relationships in the Air Force samples compared to the clerical sample. The same problems may account for failure to find aptitude effects in the remaining Air Force specialties. Another explanation concerns possible between-specialty differences in supervisors' latitude to assign tasks to enlisted personnel with different abilities.

Despite these limitations and questions, the results of the present study are viewed as sufficiently promising to warrant further exploration of the utility of occupational surveys for assessing enlisted job performance. Additional information available from the surveys would permit alternative, low-cost performance indices, as well as aptitude/experience tradeoff issues, to be examined more fully. In addition, univariate and multivariate corrections for sample curtailment on aptitude could be accomplished. Additionally, a probable confound in the present study concerning differences in actual jobs performed could be addressed. Enlisted personnel are known to change jobs in a specialty as they progress in their careers. If the job experience covariate did not provide a sufficient control, differences in positions as a function of airman abilities would have been masked. Special-purpose software developed for the Air Force occupational survey program allows job types to be distinguished (Christal, 1974). Reanalysis by job types within specialties might reveal stronger aptitude effects.

The magnitudes of aptitude and experience tradeoffs were fairly consistent with those reported by Alley and Teachout (1990). For the TOPM criteria, one year of job experience was equivalent to approximately 16 to 22 percentile points on the aptitude metric. That is, enlisted personnel with higher aptitude (16 or more points) were performing at a level about equal to their coworkers with an additional year of experience. Results for the JPM Project criteria were 10 to 15 aptitude points per year of experience. Questions remain about the utility of the TOPMs for addressing work force structure issues. Only linear effects for aptitude and experience were examined. More complex functional forms (e.g., curvilinear with aptitude and experience interaction) may be appropriate and need to be assessed. Findings would have important implicati ns to managers working downsizing questions about the costs and benefits of more or less experienced workers with different aptitude mixes in various specialties. Results would be useful for justifying decisions about the most effective combinations of personnel.

REFERENCES

Alley, W.E., & Teachout, M.S. (1990, August). Aptitude and experience trade offs on job performance. Paper presented at the American Psychological Association Convention, Boston, MA.

Christal, R.E. (1974). The United States Air Force occupational research project. (AFHRL-TR-73-75). Lackland AFB, TX: Occupational Research Division, Air Force Human Resources Laboratory.

Hedge, J.W., & Teachout, M.S. (1984). Job performance measurement: A systematic program of research and development. (AFHRL-TP-86-37). Brooks AFB, TX: Training Systems Division, Air Force Human Resources Laboratory.

Koym, K.G. (1977). Predicting job difficulty in high aptitude career ladders with standard score regression equations. (AFHRL-TR-77-26). Lackland AFB, TX: Occupation and Manpower Research Division, Air Force Human Resources Laboratory.

Nathan, B.R. (1992, April). Job progression and position appraisal: Two relevant and available criteria for test validation. Paper presented at the meeting of the Society of Industrial and Organizational Psychology, Montreal, Quebec, Canada.

Table 1

Bivariate Correlations for Performance Measures with Experience and Aptitude

				ATDPUTS		NTP		JDI	
AFS	TITLE	N	AI	Exp	Apt	Exp	Apt	Exp	Apt
113X0C	Flight Engineer	227	G	.24**	-.03	.26**	.01	.29**	-.02
271X2	Ops. Resources Mgt	728	A	.32**	-.07	.18**	-.02	.29**	-.02
302X0	Weather Equip. Spec	200	E	.37**	.16**	.10	.12*	.24**	.16*
304X1	Nav. Aids Equip. Spec	304	E	.26**	.09*	.25**	-.03	.29**	.02
306X0	Comm & Crypto Equip	580	E	.05	-.00	.13**	.06	.10*	.04
306X3	Telecomm Sys Maint Tech	460	E	.30**	.07	.34**	.04	.35**	.06
321X0	Bomb-Nav Sys Spec.	209	E	.36**	.07	.34**	-.03	.33**	.01
328X1	Avionic Nav Sys Spec	685	E	.16**	.10**	.10**	.05	.13**	.09*
361X0	Cable & Antenna Tech	263	M	.31**	.13*	.19**	.09	.30**	.11
411X2A	Missile Facilities Helper	319	E	-.04	-.08	-.05	-.14**	-.12*	-.11*
426X2	Jet Engine Mech	2011	M	.21**	.14**	.24**	.21**	.26**	.22**
426X3	Turboprop Prop Tech	458	M	.36**	-.10*	.20**	-.07	.29**	-.09
431X1	Tactical A/C Maint Spec	922	M	.26**	.11**	.02	.01	.11**	.09**
451X4B	F-15 Avionics Comp.	251	E	.09	.11*	.20**	.04	.05	.09
452X1A	F-15 Avionics Sys Helper	183	E	.18*	-.02	.28**	.03	.23**	-.00
464X0	Explosive Ord Disposal	344	M	.38**	.01	.25**	.01	.33**	-.02
491X1	Info Sys Operator	1248	G	.29**	.09**	.17**	.01	.27**	.02
491X2	Info Sys Programmer	190	G	-.02	.06	.22**	-.08	.18*	-.02
492X1	Info Sys Radio Opr	582	A	.41**	-.11**	.12**	-.03	.28**	-.06
493X0	Info Sys Control Spec	656	E	.23**	.08*	.12**	.01	.18**	.04
553X0	Engineering Asst Spec	492	G	.36**	.12**	.18**	.08*	.29**	.15**
555X0	Production Control Spec	187	G	.43**	.07	.40**	.17**	.47**	.15*
631X0	Fuel Specialist	1618	M	.34**	.04	.27**	.01	.40**	.04

* $p<.05$, ** $p<.01$

Note: AI variables = Mechanical, Administrative, General, and Electronic

ATDPUTS = Average Task Difficulty Per Unit Time Spent
NTP = Number Task Performed
JDI = Job Difficulty Index

Table 2

Multiple Correlations (R) and Results of Linear Model Comparisons (R^2 Difference) for Specialties with Significant Aptitude Contributions, Holding Experience Constant

	ATDPUTS		NTP		JDI	
AFS	R	R^2diff	R	R^2diff	R	R^2diff
328X1	.18***	.006*				
361X0	.33***	.015*				
411X2A			.14*	.018*		
426X2	.23***	.010***	.29***	.029***	.31***	.030***
431X1					.13**	.005*
491X1	.32***	.019***			.28***	.003*
553X0	.37***	.008*			.32***	.014**
555X0			.42***	.018*		
631X0	.35***	.003*			.40***	.004**

* $p<.05$, ** $p < .01$, *** $p < .001$
Note: R^2 diff is increase in Multiple Correlation when AI is added to the model

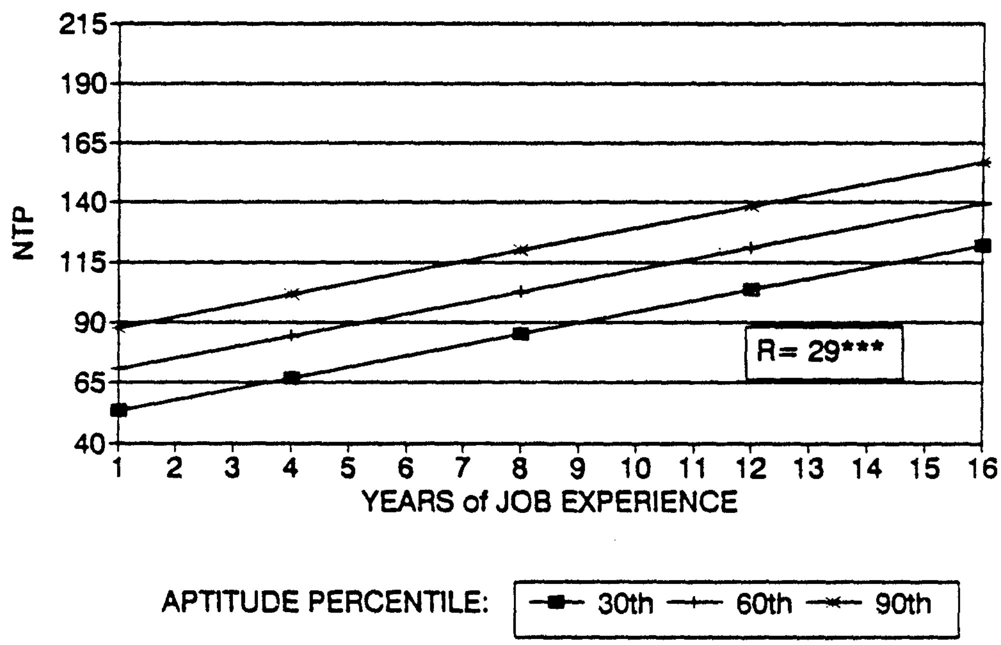


Figure 1. Expected number of tasks performed in the Jet Engine Mechanic specialty (426X2) as a function of aptitude and experience.

ENLISTED JOB CLASSIFICATION TECHNOLOGY: INITIAL DEVELOPMENT

Walter E. Driskill, Ph.D.
Johnny J. Weissmuller
Martin J. Dittmar
Metrica, Inc.

The research reported here describes a new approach to reorganizing occupational specialties. This approach utilizes the Comprehensive Occupational Data Analysis Programs (CODAP) as the analytic tool for processing information collected from job incumbents to create various classification scenarios for consideration.

Approach

Although there are several factors that impact occupational specialty structuring requirements (e. g., manpower and training requirements, promotion opportunity), this study focussed on two of them: transferability of training and opportunity to perform. The literature regarding the transfer of training suggests that an important principle to apply during restructuring consideration is to assign tasks to specialties having similar knowledge, skill, and ability requirements for the tasks being added (see, for example, Lance, Kavanaugh, & Gould, 1989, for a summary). It seems apparent, at least intuitively so, that job incumbents should be located in the workplace so that they have the opportunity to perform the tasks that are combined.

Use of CODAP job data, as well as the analytic capability of CODAP, seemed to provide a means for developing information about each of the two factors included in this study. CODAP provides a hierarchical clustering procedure that identifies tasks that are coperformed -- that is, if an incumbent performs one of the tasks in a coperformed module, the probability is high that the incumbent will perform the other tasks in the module. The coperformance modules provide information about the opportunity to perform.

Knowledge and skill (KS) requirements for performing the tasks in a module can be collected from job incumbents and processed by CODAP to compare similarity (as well as dissimilarity) of the knowledges and skills among any set of modules. Thus, the degree of transferability of knowledges and skills can be quantitatively assessed.

Use of the task coperformance module approach provides two advantages. First, modules become the unit of analyis, as opposed to more typical approaches which compare occupational specialties or those using single tasks (or a sample of tasks). These approaches can be inhibited, because, in the first instance, unique differences may be masked by consideration of the overall specialty while, in the second case, the combination of tasks becomes analytically complicated. Second, with information about KS collected at the module level, various classification scenarios can be constructed by using different combinations of modules. Restructuring decisions can then be made after assessment of the KS similarities and the impact on training and manning requirements associated with the various module combinations.

In summary, the procedure involves decomposing jobs into component task modules, describing the modules in terms of common knowledge and skill taxonomies, and regrouping the modules to form new jobs with greater knowledge and skill homogeneity.

Method

The procedure was applied to two flightline maintenance occupational specialties, each specialty having three shredouts corresponding to different aircraft avionic subsystems. The only difference between the two specialties was aircraft designation, each specialty referring to a different aircraft. (Note: Previously, the work was structured around the avionic subsystems, with type of aircraft being used as shredouts).

Occupational data collected by the U.S. Air Force Occupational Measurement Squadron was analyzed to identify the task coperformance modules for each specialty. The coperformance modules, as one might expect, reflected combinations of tasks performed by job incumbents having a specific avionics-system shred. A total of 23 coperformance modules were identified for one specialty, 46 for the other.

Three taxonomies were developed: 1) electronics fundamentals knowledges; 2) aircraft avionic subsystem knowledges; and 3) maintenance skills. In defining these three taxonnomies, effort was directed at identification of sets of taxonomy elements that were discrete, interpretable by job incumbents, and of sufficient scope to encompass job activities associated with the specialties. The electronics fundamentals were obtained from the Electronics Principles Inventory. Aircraft subsystem knowledge was developed from maintenance data for each of the two aircraft. Skills were identified from the "action taken" codes used for reporting aircraft maintenance.

Mail survey methods were employed. Each job incumbent was provided a survey booklet containing a set of task coperformance modules which the incumbent rated using 9-point relative importance scale. Incumbents rated the importance of fundamentals, system knowledges, and skills separately.

Rater reliabilities for fundamentals and skills were calculated using the CODAP GRPREL program. Interrater reliabilities for a single rater and for each composite of N raters were calculated for each coperformance module. The median single-rater reliabilities ranged from .13 to .60 for fundamentals; from .46 to .62 for skills. Composite reliabilities ranged from .63 to .94 for fundamentals; from .94 to ..97 for skills.

Split-half reliabilities were computed for each task module for system knowledges. Pearson correlations ranged from a low of .32 to .99. The median correlation across forms and modules was .88.

Mean values for the ratings of the elements from the taxonomies were computed for each task module. Task modules were hierarchically clustered, using the CODAP OVERLAP program. This procedure produced clusters of task modules based on their similarity of fundamental knowledges, system knowledges, and skill requirements.

Results

Given that a major goal of this effort was to determine if valid information about knowledge and skill requirements can be elicited from job incumbents via mail survey, moderate success was achieved. As expected, systems knowledges cleanly differentiated specialty shredouts. The results indicated that system knowledge requirements for the same shredouts across specialties are more similar than the requirements across the shredouts within a specialty.

Although some differences among shredouts were noted with respect to electronics fundamentals knowledges, the predominant requirements across shreds were similar: knowledge of multimeters, soldering or solderless connectors, and direct-alternating current. Similarly, few differences were noted either among shredouts or specialties with regard to skill requirements.

Conclusions

Results of the study tend to demonstrate the merit of eliciting knowledge and skill requirements for performing the tasks of a task coperformance module from job incumbents using a mail survey approach. Information obtained from the systems knowledge taxonomy provided better differentiation among task modules than did either electronic fundamentals knowledges or skill requirements.

The methodology has the potential for providing two important kinds of information useful for considering classification restructuring issues. It provides a convenient, flexible unit of analysis for considering various combinations of jobs in terms of the opportunity job incumbents have for performing in new or reorganized specialties, as well as in terms of the quantitative similarity of knowledges and skills required to perform in the new specialties.

Reference

Lance, C.E., Kavanagh, M.J., & Gould, R.B. (1989). Development and convergent validation of cross-job ease-of-movement indices. Paper presented at the annual convention of the American Psychological Association, New Orleans, LA.

Development of a Career Field Management Plan for the Electronic Computer and Switching Systems Specialty

Captain Terri Coccia
USAF Occupational Measurement Squadron

Jimmy L. Mitchell & J. R. Knight
McDonnell Douglas Corporation

Lt Rory C. Shrum
Human Resources Directorate
Armstrong Laboratory

Abstract

A joint study of the Electronic Computer and Switching Systems specialty was undertaken by the USAF Occupational Measurement Squadron and the Technical Training Research Division of the Armstrong Laboratory, as reported at the 1991 Military Testing Association Conference. The parallel occupational analysis and Training Decision System (TDS) studies were both aimed at assisting the AFS 305X4 Functional Manager in making critical decisions. The TDS results were used to model the current state of the specialty as well as several alternative structures proposed by an Subject-Matter Expert panel which met in May 1992. Data from both efforts will be briefed to a Utilization and Training Workshop (U&TW) for the specialty in late 1992, and the information will be used by U&TW participants to make decisions about the future course of training in the specialty. Some of the data from the occupational analysis and the TDS study were also used to draft sections of a Career Field Management Plan (CFMP), a newly-required planning document which highlights expected changes in manpower, personnel, and training programs over the next five years. A CFMP will be required of all Air Force specialties in the near future; the present study has demonstrated that both occupational analysis and TDS data can be useful in CFMP development.

Introduction

The USAF Occupational Measurement Squadron and the Technical Training Research Division of the Armstrong Laboratory, Human Resources Directorate, have been cooperating in an operational test of the Training Decisions System (TDS). The TDS is a computer-based training requirements planning and decision support system, and many aspects of the project have been briefed to earlier Military Testing Association conferences (Bennett, et al., 1991; Ruck, 1989; Mitchell, et al., 1991). We have also published a series of associated technical reports and papers (Vaughan, et al., 1989; Mitchell, et al., 1992). TDS can help manpower, personnel, and training (MPT) and functional managers to visualize and understand dynamic job flows and available training programs of the occupation under consideration.

This system also assists managers and analysts in evaluating MPT policy options in terms of costs and training capacities of representative units, and in conducting trade-off analyses between various formal training programs and OJT. The TDS supports Air Force managers in making decisions as to the <u>what</u>, <u>where</u>, and <u>when</u> of the technical training (including OJT) required for an occupation (Ruck, 1989). It includes procedures for developing data bases and modeling the dynamic flow of people through jobs and through both formal training and OJT. Furthermore, the system includes modeling and optimization capabilities which provide estimates of training quantities, costs and capacities for both formal training and on-the-job training (Vaughan, et al., 1989).

Our plans for the cooperative project on the Electronic Computer and Switching Systems (AFS 305X4) specialty were briefed to this conference last year (Mitchell, Buckenmyer, Huguley, and Knight, 1991). We had hoped to have the work completed by this time, including the final step of briefing the AFS 305X4 Utilization and Training Workshop (U&TW).

But reality does not always proceed as planned. Indeed, the project encountered some delays almost immediately after our report at last year's MTA meeting.

Field Administration of the 305X4 Job Inventory

The administration of the AFS 305X4 Job Inventory was scheduled to close in November 1991, but not enough returns had been received by that date to insure a satisfactory sample of the specialty. The job inventory used in this field administration included a four page insert which was the TDS Job and Training History Survey (J&THS). By administering these instruments jointly, we avoided having two surveys in the field at the same time, and also eliminated the respondents background section from the J&THS. By using the same booklet control number, we could match up the Job Inventory and the J&THS data files in the computer, and thus could make use of the extended background section for either analysis.

By late November, the occupational analyst had concluded that the returned surveys did not include some of the critical bases which needed to be in the final sample. A second set of booklets was sent to several such bases for a supplemental administration; this delayed the closeout of the study until late in January 1992.

AFS 305X4 Data Analyses

The AFS 305X4 occupational survey data were scanned and loaded to the USAFOMSq IBM mainframe computer; this was one of the first studies to be accomplished solely on that system. For the TDS study we wanted to utilize some of the still-experimental advanced task clustering programs (Mitchell, Phalen, Haynes, & Hand 1989) to expedite our analysis; these programs were not yet available on the IBM system. Thus, it was decided to do a parallel analysis on the Armstrong Laboratory UNISYS since the experimental programs could be utilized on that system. Thus, parallel analyses of the AFS 305X4 occupational survey data on the two different systems were accomplished.

For the TDS study, advanced task clustering programs were used to develop a set of Task Modules for the specialty. The automated MODTYP results were interpreted and given preliminary titles pending a review by subject matter experts (SMEs). One AFS 305X4 SME assigned to HQ ATC at Randolph AFB, TX provided an initial screening of the TMs including assignment of unclustered tasks to the most appropriate TM. At the same time, the Armstrong Laboratory coordinated with the Air Staff functional manager for this specialty, to arrange for a representative panel of SMEs who would be able to review and validate the Task Modules (TMs) as well as provide much of the other data needed to develop an AFS 305X4 TDS data base.

SME Review Panel

A review panel of SMEs representing most of the functional areas of the Electronic Computer and Switching Systems specialty was convened at Brooks AFB, TX in early May 1992. They represented three technical training centers, most major commands, and several other specialized units. The group was chaired by the Air Staff functional manager, who noted that this meeting was held to prepare for the U&TW which would follow the publication of the occupational survey report. The meeting was funded in part through Armstrong Laboratory research and development (R&D) funds and also by the functional manager. This joint funding of the SME panel demonstrates how operational and R&D needs can both be satisfied in a cooperative project.

The SME panel reviewed and validated the set of 73 Task Modules (TMs) with which to characterize the specialty (see Figure 1); however, their review also revealed that several of these TMs involved equipment or systems which being phased out in the near future. The SME panel also titled each TM and provided data on representative sites (bases and units) where various systems were maintained. In addition, the job groups identified by both the occupational analyst and TDS analyst were reviewed and named (see Figure 2); by having an agreed upon set of clearly recognizable job types, both studies benefited and better communication with AFS incumbents was assured.

1 GENERIC MAINTENANCE TASKS
2 PERFORM GENERAL MAINTENANCE
3 MAINTAIN DISC SYSTEMS
4 MAINTAIN PRINTERS

. .

. .

10 BENCH CHECK ASSEMBLIES
11 TROUBLESHOOT/REPAIR TO COMPONENT LEVEL
12 MAINTAIN BUFFERS, CONTROLLERS OR OTHER INTERFACES
13 MAINTAIN RAPID ACCESS MAINTENANCE MONITOR (RAAM) INTERFACE CIRCUITS
14 MAINTAIN MULTIPLEXERS

. .

. .

31 MAINTAIN MOBILE COMPUTERS OR SWITCHING SYSTEMS
32 MAINTAIN TAPE PUNCH/READER EQUIPMENT
33 MAINTAIN LIGHT PENS
34 MAINTAIN ANALYTICAL PHOTOGRAMMETRIC POSITION SYSTEMS (APPS)
35 MAINTAIN PLOTTERS

. .

. .

71 MAINTAIN AUDIO SYSTEMS
72 MAINTAIN ELECTRONIC TYPEWRITERS
73 CAREER DEVELOPMENT COURSE

Figure 1. Examples of Electronic Computer and Switching Systems (AFS 305X4) Task Modules

General Computer Maintenance Technician (E-shred; N = 422)
SACDIN and General Maintenance (L-shred; N = 134)
Mobile Systems Maintenance (N = 157)
Computer Maintenance - Cheyenne Mt (R-shred; N = 16)
Joint Surveillance Systems (JSS) Maintenance (M-shred; N = 58)
Small Computer Maintenance (N = 14)
AWACS Component Repair (N = 23)
AWACS On Aircraft Systems (T-shred; N = 53)
DDS Operator/Maintainer (N = 31)
Secure Command and Control System (Secure Voice; N = 16)
General Computer Maintenance NCOICs/Shift Supervisors (N = 20)
Work Center Supervisors (N = 121)
Crew Chiefs/NCOICs (N = 47)
Work Center Supply Monitor (N = 11)
Technical Instructors (N = 58)
Small Computer Maintenance/Requirements (N = 12)
Instructor Supervisors (N = 10)
Quality Control (N = 30)
Logistical Support (N = 14)
Program Management (N = 23)
Job Control (N = 43)
Staff NCOs (Computer Systems Maintenance Overhead; N = 16)

Figure 2. Electronic Computer and Switching Systems (AFS 305X4) Jobs

Additionally, the panel provided a variety of other types of information needed for the TDS study of their specialty. Such data included allocation data with which to construct learning curves for each TM, the training resources required for training each TM, representative sites, and projected or proposed changes in the specialty expected over the next five years. They also identified key NCOs at each base to serve as contacts for the collection of additional data needed to complete the TDS data base.

The Air Staff functional manager also provided a manpower data base (current and projected FY95 authorizations by unit) which could be used to track recent changes in units and expected future changes in manpower levels. This data base proved to be exceptionally worthwhile since it included recent organizational changes (MAJCOMs, redesignated units, etc.). For example, it identified certain units which, while manned during the Job Inventory field administration, would cease to exist prior to the U&TW (such as Bergstrom AFB, TX, and other bases on the base closure list). This permitted us to build a more up to date model of the current specialty than would otherwise have been possible.

Overall, the AFS 305X4 SME Panel proved to be a very successful conference. While some additional data had to be collected by field survey and telephone contacts (for a few functional areas not represented), the Panel was able to provide a very solid foundation for modeling this very complex specialty. The representatives also benefited through their interaction with the Air Staff functional manager and among themselves, as well as getting a preliminary perspective on the results of the occupational survey and the possible outcomes from the TDS study. Thus, the conference demonstrated the value of this type of meeting, not just for the TDS R&D but also for the occupational analyst and the SMEs as well. Further, the objectives of the Air Staff functional manager were met in terms of the preliminary discussion of many of the issues which will be of concern at the U&TW later this fall.

Interchange of Data

Based in part on the successful exchange of information at the SME Panel meeting, further interchange of information has continued to occur between the occupational analysis and TDS projects. We have transported additional data files and completed products electronically between the OMSq and AL computers, and we have exchanged ideas and information between analysts. The TM-level job descriptions for the agreed-upon job groups were provided to the occupational analyst. In turn, she made available her matching of new job types with those from the last OSR for TDS use in interpreting J&THS results. For the TDS study we have also examined other data bases, and have made use of historical Uniform Airman Record (UAR) data tapes stored at AL, as well as Air Training Command records and other files, to estimate entries into the career field, attrition from courses and the specialty, and other types of information needed in the TDS model of the AFS.

TM-Level Task Factor Data

In the interchange of AFS 305X4 information, we were also asked if we could display Task Difficulty (TD) and Training Emphasis (TE) data in conjunction with the TMs of the specialty. The particular issue involved was that the OMSq task analysis unit was trying to prioritize among the areas of this specialty for a detailed task analysis effort; there was not sufficient time or resources to analyze all the tasks within the AFS.

By transfering the task factor data files from the OMSq computer to the UNISYS at AL, we were able to generate a Task Within TM listing which also displayed TD and TE ratings, as well as the module-level averages for these task factors (see Figure 3). This new product proved very useful, particularly since the TMs had been reviewed and validated by the SME Panel earlier in the year. In their review, some TMs with low percent performing (total sample) were identified as obsolete systems which were going out of the inventory; this tended to be confirmed by the very low TE ratings received.

With this added information, it was very easy to eliminate these TMs from any consideration for the detailed task analysis project thus saving a substantial amount of time and energy. The remaining TMs were also prioritized, in terms of a composite judgement involving TE, TD, total sample percent performing, and total sample percent time spent, to provide for further savings of time and effort in detailed task analysis.

These same low-performance and obsolete equipment TMs can also be eliminated from the TDS model of the specialty since they are no longer relevant work for which AFS incumbents must be trained. By elmininating these TMs from the model, we can provide a more realistic view of what the AFS is currently and what it is expected to be in the future. In addition, TDS data for these TMs is difficult if not impossible to collect.

PRTMOD Total sample (AFS 305X4, n = 1804) Page 3

TM	Module Title	No. of Tasks	Percent Time Spent Sum	Cum	Avg	Avg PMP	TD *	TE **
63	GENERAL ADMINISTRATIVE ACTIVITIES	7	3.48	3.48	.50	27.20	6.1	1.1
A 39	Participate in meetings, such as staff meetings, briefings, conferences or workshops		.83	.83		44.07	5.8	.9
A 30	Draft or write correspondence		.68	1.51		33.20	6.6	1.7
B 49	Compile information for reports or staff studies		.49	2.00		22.06	6.2	1.0
E 282	Type correspondence, forms, or reports		.47	2.47		29.27	5.7	1.4
C 138	Review correspondence		.42	2.88		23.23	6.2	.9
B 50	Conduct briefings		.34	3.22		22.34	6.1	1.0
A 42	Plan briefings		.26	3.48		16.24	6.1	.8
.								
.								
04	MAINTAIN PRINTERS	4	1.66	19.15	.41	44.29	6.3	4.6
H 411	Perform PMIs on printers		.58	.58		52.00	6.1	4.7
J 532	Isolate malfunctions within printers to cards or subassemblies		.38	.95		43.02	6.7	4.7
I 440	Align printers		.37	1.32		41.08	6.6	5.0
K 603	Remove or replace printer subassemblies		.34	1.66		41.08	5.7	3.9
.								
.								
33	MAINTAIN LIGHT PENS	2	.05	99.15	.02	4.05	3.8	1.6
K 581	Remove or replace light pens		.03	.03		4.71	3.5	1.5
J 460	Isolate malfunctions in light pens		.02	.05		3.38	4.0	1.7
.								
.								

* Average TD = 5.0; S.D. = 1.00
** Average TE = 2.0; S.D. = 1.35

Figure 3. Example Task & TM-Level Task Factor Data (AFS 305X4)

This link up of TE and TD data with TMs, and computation of average values at the TM level has some interesting implications for both TDS and for refinement of the task clustering methodologies. Task modules may be the more appropriate level at which some types of information about a specialty should be gathered. There have been some significant studies reported recently where module level ratings were evaluated as indicators of occupational requirements and to prioritize training (Moon, Driskill, Weissmuller, Stayer, Fisher, & Kirsh, 1991). Additional research in this area is greatly needed.

Utilization and Training Workshop

The next major event in the Electronic Computer and Switching Systems (AFS 305X4) project will be the U&TW, which will be convened by the Air Staff functional manager, probably in late November or early December of this year. At that conference, both the results of the occupational survey and the model of the specialty developed in the TDS will be briefed to participants.

Several preliminary alternative utilization and training patterns are being modeled as requested by the functional manager at the SME Panel meeting in May. In addition, other alternatives can be modeled as they are suggested by participants at the U&TW.

We have every reason to believe that this will be a very successful meeting, and that both the OSR and the TDS data will be useful to U&TW participants as they plan out the expected development of this specialty for the next few years. Indeed, we hope to continue to work with this community to help with their decision making and documenting their planning.

Career Field Training Management Plan

One of the expected outcomes of the U&TW will be the development of a career field training management plan (CFTMP) which will document the decisions made by the U&TW as to how training for the specialty should be done. This plan, which is expected to be mandated for all specialties over the next two years, will describe and characterize the AFS, will examine expected changes in the specialty (in terms of manpower changes, organizational restructurings, and expected changes in equipment and systems), and will plot out how training programs (including on-the-job training programs) will need to be modified to meet the new requirements.

We hope to play a major role in the development of such a CFTMP in that the OSR data on AFS jobs, AFR 39-1 specialty descriptions, and initial skills training programs can be useful in determining needed AFS changes. The TDS model of this specialty can also be used in terms of assessing the impacts of the changes on specialty training programs and typical unit training capacities. U&TW participants may be able to use this kind of information in terms of evaluating and deciding among several restructuring options. In addition, some of the OSR and TDS data may be useful directly in terms of drafting the CFTMP; indeed, we are hoping to draft some portions of this plan for review and possible adoption by the U&TW and Air Staff functional manager. Since only one or two CFTMPs have been developed to date, and since there is not absolute standard for what CFTMPs should be, there is a possibility that whatever CFTMP is developed and adopted for this specialty (AFS 305X4) may indeed become the prototype for what a CFTMP should be in the future.

References Cited

Bennett, W.R., Ruck, H.W., Vaughan, D.S., & Mitchell, J.L. (1991, October). Training and organizational development in military organizations: Implications for diagnosis and needs assessment. *Proceedings of the 33d Annual Conference of the Military Testing Association.* San Antonio, TX: Human Resources Directorate, Armstrong Laboratory and the USAF Occupational Measurement Squadron.

Mitchell, J.L., Buckenmyer, D.V., Huguley, H., & Knight, J.R. (1991, October). Development of an integrated data base for the electronic computer and switching systems specialty (AFS 305X4). *Proceedings of the 33d Annual Conference of the Military Testing Association.* San Antonio, TX: Human Resources Directorate, Armstrong Laboratory and the USAF Occupational Measurement Squadron.

Mitchell, J.L., Phalen, W.J., Haynes, W.R., & Hand, D.K. (1989, October). *Operational testing of ASCII CODAP job and task clustering methodologies* (AFHRL-TP-88-74). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Mitchell, J.L., Vaughan, D.S., Knight, J.R., Rueter, F.H., Fast, J., & Haynes, W. R. (1992, June). *Training decisions technology analysis* (AL-TP-1992-0026). Final Technical Report. Brooks AFB, TX: Technical Training Research Division, Armstrong Laboratory, Human Resources Directorate.

Moon, R.A.J., Driskill, W.E., Weissmuller, J.J., Strayer, S.J., Fisher, G.P., & Kirsh, M. (1991, October). Using task co-performace modules to define job requirements. *Proceedings of the 33d Annual Conference of the Military Testing Association.* San Antonio, TX: Human Resources Directorate, Armstrong Laboratory and the USAF Occupational Measurement Squadron.

Ruck, H.W. (1989, November). Symposium: The Air Force training decisions system - Modeling management and policy issues for efficient training. *Proceedings of the 31st Annual Conference of the Military Testing Association.* San Antonio, TX: Air Force Human Resources Laboratory and the USAF Occupational Measurement Center.

Vaughan, D.S., Mitchell, J.L., Yadrick, R.M., Perrin, B.M., Knight, J.R., Eschenbrenner, A.J., Rueter, F.H., & Feldsott, S. (1989, June). *Research and Development of the Training Decisions System* (AFHRL-TR-88-50). Brooks AFB, TX: Training Systems Division, Air Force Human Resources Laboratory.

MEASUREMENT EQUIVALENCE AND SCORE EQUATING FOR EXPERIMENTAL AND OPERATIONAL BASIC ATTRIBUTES TEST

Thomas R. Carretta and Malcolm James Ree

Armstrong Laboratory Human Resources Directorate
Brooks Air Force Base, Texas

The Air Force plans to phase-in a new pilot candidate selection system in the near future. Scores from the Air Force Officer Qualifying Test (AFOQT, see Skinner & Ree, 1987), Basic Attributes Test (BAT, see Carretta, 1989, 1992), and biographical information will be used to compute a pilot candidate selection composite score. The AFOQT, though highly *g* loaded (Earles & Ree, 1991) is a paper-and-pencil test battery which assesses five factorial ability domains: verbal, quantitative, spatial, perceptual speed, and aircrew interest/aptitude (Skinner & Ree, 1987). Fourteen of the 16 AFOQT subtests contribute to the Pilot and Navigator-Technical composites which are used for pilot candidate selection (U.S. Air Force, 1983). The BAT is a computer-based multiple aptitude test battery which measures individual differences in psychomotor skills, cognitive abilities, personality, and attitudes toward risk (Carretta, 1987, 1989). Several studies have demonstrated the incremental validity of BAT scores for predicting flying training performance (Carretta, 1989, Kantor & Carretta, 1988) and its robustness under cross-validation (Carretta, 1990). Additional pilot candidate selection factors include medical fitness, academic performance, aptitude test scores, and previous flying experience.

The new pilot candidate selection composite, known as the Pilot Candidate Selection Method or PCSM, was developed using an experimental form of BAT hardware and software. The experimental BAT systems are too few in number and difficult to maintain to support full-scale operational testing of Air Force pilot training applicants. The Air Force selects pilot candidates from several sources including the Air Force Academy (AFA), Officer Training School (OTS), Reserve Officer Training Corps (ROTC), and active duty. The Air Force plans to field over 150 test systems at ROTC detachments, Military Enlistment Processing Stations, and other locations (i.e., AFA). The purpose of this study was to equate the experimental and production tests. The pre-implementation equating was necessary (a) to determine whether the production prototype tests measured the same psychological constructs as the experimental tests (i.e., construct validity), (b) to compare the score distributions of the two tests, and (c) to develop equating tables for placing scores from operational tests on the experimental test metric. Equivalency was required so that scores from the operational tests could be used in proposed Air Force pilot candidate selection equations developed on the experimental BAT system.

METHOD

Subjects

The subjects were 2,034 Air Force recruits with a median age of 19 years and were mostly White (84%), male (69%), and high school graduate or better (99%). All subjects were selected for training, in large part, on the basis of their Armed Services Vocational Aptitude Battery (ASVAB, see Earles & Ree, in press; Ree, Mullins, Mathews, & Massey, 1982) scores, and educational achievements.

Apparatus

The BAT apparatus consisted of a microcomputer and monitor built into a carrel with a glare shield and side panels designed to eliminate distractions. Each subject responded to the tests by using, individually or in combination, a dual-axis control stick on the right side of the apparatus, a single axis control stick on the left side, and a specialized keypad in the center. The keypad included the number keys 0 to 9, an ENABLE key in the center, and a bottom row with YES and NO keys and two others labeled S/L (for Same/Left responses) and D/R (for Different/Right responses).

Measures

The BAT was designed to measure individual differences in psychomotor skills, cognitive abilities, personality, and attitudes toward risk. The type of scores generated from these tests included tracking error/difficulty, response speed, response accuracy, and response choice. A brief description of the tests follows and a more detailed description was provided by Carretta (1989, 1990).

Two-Hand Coordination. This was a rotary pursuit task (Fleishman, 1964). An airplane (target) moved in a fixed, elliptical pattern at a varying rate. The subject controlled the horizontal and vertical movement of a "gunsight" using the right (horizontal) and left (vertical) control sticks. The subject's task was to keep the gunsight on the target.

Complex Coordination. This stick and rudder, compensatory tracking task was an example of multilimb coordination (Fleishman, 1964). The dual-axis right control stick was used to control the horizontal and vertical movement of a cursor. The left control stick was used to control the left-right movement of a "rudder bar" at the base of the screen. The subject's task was to maintain the cursor (against a constant horizontal and vertical rate bias) centered on a large cross fixed at the center of the screen, while simultaneously centering the rudder bar at the base of the screen (also, against a constant rate bias).

Encoding Speed. This task assessed verbal processing efficiency and was adapted from a paradigm proposed by Posner and Mitchell (1967). The subject was presented simultaneously with two letters and required to make a same-different judgment about the letter pair. The complexity of the decision rule increased throughout the task.

Mental Rotation. This was a variation of a spatial transformation task (Shepard & Metzler, 1971). The subject was presented sequentially with two letters and required to make a same-different judgment. The letter pair consisted of either same or mirror images and the letters were either in the same orientation or rotated in relation to each other. A correct "different" judgment is associated with a mirror image pair and is not dependent on the relative rotation of the two letters.

Item Recognition. This measure of short-term memory was based on a task proposed by Sternberg (1966). A string of 1 to 6 digits was presented on the screen. The string was then removed, and after a brief delay, replaced by a single digit. The subject was instructed to remember the digit string and indicate whether the single digit was one of those presented in the digit string.

Time Sharing. This test provided a measure of time sharing performance (North & Gopher, 1976). In the first 10 minutes of this test, the subject was required to keep a randomly moving "gunsight" on an airplane (target) using the right-hand control stick. In the next six minutes, the subject had to repeat the tracking task and simultaneously cancel digits which appeared at random intervals and locations on the screen. Digit cancellation was timed and consisted of pressing the same digit on the numeric keypad. The final three minutes of the test consisted of tracking only. Tracking difficulty was varied by increasing or decreasing the control stick sensitivity as a function of the tracking error.

Self-Crediting Word Knowledge. This was a vocabulary test where the subject was required to make predictions about his or her performance (i.e., self-confidence) prior to each group of items (Mullins, 1962). The subject had to predict if they would know the meaning of the next word. Items were multiple choice and used synonyms.

Activities Interest Inventory. This test was designed to measure the subject's attitudes toward risk-taking (Mullins, 1962). The subject was presented with 81 pairs of activities and was asked to choose between each pair. The activity pairs forced the subject to choose between activities that differed as to degree of threat (sometimes subtly, sometimes not).

Equating

Equating is the process by which test scores are made comparable by placing both on the same metric. Angoff (1974) provided the commonly accepted definition of equating.

A commonly accepted definition of equivalent scores is: Two scores, one on Form X and the other on Form Y (where X and Y measure the same function with the same degree of reliability), may be considered equivalent if their corresponding percentile ranks in any given group are equal (p. 563).

There are two usual types of equating derived from the commonly accepted definition. These are equipercentile equating and linear equating. If the two tests have the same distributional shape as demonstrated by equal measures of skewness and kurtosis, then linear equating may be conducted. Linear equating sets z-scores equal in both distributions and z-score equivalent raw scores are thereby equal. Equipercentile equating is the non-linear analogue (Angoff, 1971) which produces an area-transform on both distributions to make each rectilinear (i.e., percentiles). As percentiles are set equal (i.e., 95 on X to 95 on Y, 90 on X to 90 on Y, and so on) raw scores are set equal and thereby equated.

Because linear equating uses only two estimates, mean and standard deviation, it is desirable to use it whenever possible. Equipercentile equating uses many estimates, one for each percentile and is more sample dependent and less stable. Also, equipercentile equating requires smoothing of the resultant equating functions. This was accomplished by first, second, and third order polynomial regressions which were constrained to be monotonic. These regression coefficients add to the number of estimates and the instability of equipercentile equating. When operational and experimental tests had approximately the same skewness and kurtosis, linear equating was used; otherwise, equipercentile equating was used.

Procedures

The BAT was administered on the 11th day of basic military training and the subjects were told that the test scores were being collected for experimental purposes only. In accordance with the Privacy Act, all subjects were given the opportunity to decline participation, none did.

The eight-test BAT battery, including breaks between tests to reduce physical and mental fatigue, requires about 2-hours, 40 minutes to complete. As subjects needed to test twice on the battery to estimate retest reliability, total testing time for the eight-test battery would have been about 5.5 hours. To accommodate time constraints and minimize subject fatigue effects, the eight tests were divided into two, four-test batteries.

Battery A consisted of a test-battery introduction and the first four tests described earlier (Two-Hand Coordination, Complex Coordination, Encoding Speed, and Mental Rotation). Battery B consisted of the introduction and last four tests (Item Recognition, Time Sharing, Self-Crediting Word Knowledge, and Activities Interest Inventory). Suojects were assigned to one of eight test conditions based on test battery content (Battery A or Battery B) and test apparatus (experimental twice, experimental-operational, operational-experimental, or operational twice). Reliability was estimated by testing twice on either the experimental or operational battery. The research design used randomly equivalent groups (Campbell & Stanley, 1966).

Analyses

Analyses were performed at the test score level and included descriptive statistics of the scores, t-tests between the means, F-tests between the variances, chi-square tests of the distributions, and correlations between the scores. Correlations between tests from experimental and operational apparatuses were adjusted for unreliability using the formula (Stanley, 1971):

$$\text{Adjusted } r_{eo} = \frac{r_{eo}}{\sqrt{r_{ee} r_{oo}}}$$

Where

r_{eo} = the correlation between experimental and operational BAT scores (combining data from the exp-op and op-exp conditions)

r_{ee} = the test-retest reliability of the experimental BAT (from the exp-exp condition)

r_{oo} = the test-retest reliability of the operational BAT (from the op-op condition)

The minimum acceptable adjusted correlation was set at .85. The measure of equating goodness (Braun &

Holland, 1982) is how well the new test, *after equating*, has replicated the distributional shape of the old test. This was investigated and evaluated in this study, using the chi-square goodness-of-fit statistics.

To determine equivalence of means, variances, and distributional shapes, t-tests, F-tests, and chi-square tests were computed, respectively. The $p < .05$ Type I error rate was used on all observed data. No corrected correlations were tested.

Several of the test scores are measured in very small and precise increments (e.g., tracking error in pixels, response times in milliseconds). Each subject could potentially receive a unique value for these scores leading to several hundred values which would make it very difficult to equate the forms. To reduce the number of unique values, score intervals dependent on the units of the measure, were created before performing the equatings.

RESULTS

Descriptive Measures

Examination of descriptive statistics revealed several subjects with either missing or extreme values for one or more scores. Subjects with missing data were removed from the samples. Scores of extreme values (i.e., more than three standard deviations from the mean) were reviewed on a subject-by-subject basis. Those whose score profile suggested poor performance due to carelessness or lack of motivation were removed (Ree, Mathews, Mullins, & Massey, 1982).

Comparisons of "first test" scores from the experimental and operational batteries revealed several differences in score means (t-tests), variances (F-tests), and distribution shape (chi-square tests). When mean differences occurred, with the exception of Two-Hand Coordination vertical tracking error, the test on the operational system was easier.

Adjusted correlations demonstrated an acceptable level of agreement between scores generated from the experimental and operational systems. Measurement equivalency was confirmed. The two scores with adjusted correlations below the target value of .85 (Item Recognition - % correct, .839 and Activities Interest Inventory - Avg RT, .829) were above the lower limit of the 95% confidence interval around the .85 target correlation.

Equating

Subjects tested on the experimental apparatus during the first administration (experimental-experimental or experimental-operational) served as the anchor group during equating. This placed scores from the operational tests on the experimental form metric.

Several equating methods were compared to determine the most appropriate method for each score. Selection was based on the similarity of the distributions for the operational scores and the corresponding scores from the anchor (i.e., experimental) test. The standard error of estimate (SEE) from the polynomial regression smoothing was used as a goodness-of-fit measure to determine which smoothing method would be chosen for each of the equipercentile equatings. If one method resulted in a substantially smaller SEE than another, the method with the smaller SEE was chosen. When two smoothing methods did not differ greatly, the lower order polynomial regression equation was chosen because it involved fewer parameter estimates and would be more stable.

Results were mixed. For the psychomotor tests (Two-Hand Coordination and Complex Coordination), the cubic polynomial smoothing method produced the best results for four of

the five tracking error scores. Linear equating was chosen for the vertical tracking score from Two-Hand Coordination. Quadratic polynomial smoothing was chosen for the Psychomotor composite which is a unit weighted composite of the psychomotor scores.

Linear equating was considered the best method for the remaining scores, with the exception of Activities Interest Inventory response time (equipercentile-quadratic). Equating tables were developed to convert the scores from the operational test to the raw score based on the experimental test metric.

DISCUSSION

The goals of this study were to determine the measurement equivalence (i.e., construct validity) of the operational tests and to derive scores on the operational tests that were comparable to scores on the experimental tests. Adjusted correlations between the experimental and operational tests were acceptable (i.e., the two forms of the tests measured the same construct). Differences in the shapes of the score distributions made it necessary to develop equating tables to convert scores from the operational tests to the experimental metric. Doing so made it appropriate to use scores from the operational tests in a pilot candidate selection model developed using the experimental tests. The differences in the scores between the experimental and operational apparatus were noteworthy. Despite assurances that the operational hardware and software would emulate the experimental hardware and software exactly, substantial differences were observed in scores. No assumptions of equivalence should be made. Evaluation of the effects of hardware and software on test performance is mandatory.

Use of the operational test is scheduled for the near future. Test scores will be renormed to the operational metric and a new pilot candidate selection equation will be developed after a sufficient number of applicants have been tested and have completed training.

REFERENCES

Angoff, W.H. (1971). Scales, norms, and equivalent scores. In R.L. Thorndike (Ed.) *Educational Measurement* (2nd ed.). Washington, DC: American Council on Education.

Braun, H. I., & Holland, P.W. (1982). Observed score test equating: A mathematical analysis of some ETS equating procedures. In P.W. Holland & D.W. Rubin (Eds.), *Test Equating*. New York: Academic Press.

Campbell, D.T., & Stanley, J.C (1966). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.

Carretta, T.R. (1987). *Basic Attributes Test (BAT) system: Development of an automated test battery for pilot selection* (AFHRL-TR-87-9, AD-A185-649). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Carretta, T.R. (1989). USAF pilot selection and classification systems. *Aviation Space and Environmental Medicine*, *60*, 46-49.

Carretta, T.R. (1990). Cross-validation of experimental USAF pilot training performance models. *Military Psychology*, *2*, 257-264.

Carretta, T.R. (1992). Understanding the relationship between selection factors and pilot training performance: Does the criterion make a difference? *International Journal of Aviation Psychology, 2, 95-105.*.

Earles, J.A., & Ree, M.J. (in press). The predictive validity of the ASVAB for training grades. *Educational and Psychological Measurement.*

Earles, J.A., & Ree, M.J. (1991). *Air Force Officers Qualifying Test (AFOQT): Estimating the general ability component* (AL-TP-1991-0039). Armstrong Laboratory, Human Resources Directorate, Manpower and Personnel Research Division: Brooks AFB TX.

Fleishman, E.A. (1964). *The structure and measurement of physical fitness.* Englewood Cliffs, NJ: Prentice Hall.

Kantor, J.E. & Carretta, T.R. (1988). Aircrew selection systems. *Aviation Space and Environmental Medicine, 59,* Supplement, A32-A38.

Mullins, C.J. (1962). *Objective tests of self-confidence* (PRL-TM-62-66). Lackland AFB, TX: Selection and Classification Branch, Personnel Research Laboratory.

North, R.A., & Gopher, D. (1976). Measures of attention as predictors of flight performance. *Human Factors, 18,* 1-14.

Posner, M.I., & Mitchell, R.F. (1967). Chronometric analyses of classification. *Psychological Review, 74,* 392-409.

Ree, M.J., Mathews, J.J., Mullins, C.J., & Massey, R.H. (1982). *Calibration of Armed Services Vocational Aptitude Battery Forms 8, 9, and 10* (AFHRL-TR-81-49). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Ree, M.J., Mullins, C.J., Mathews, J.J., & Massey, R.H. (1982). *Armed Services Vocational Aptitude Battery: Item and factor analysis of forms 8, 9, and 10* (AFHRL-TR-81-55). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Shepard, R.N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 171,* 701-703

Skinner, J. & Ree, M.J. (1987). *Air Force Officer Qualifying Test (AFOQT): Item and factor analysis of form O* (AFHRL-TR-86-68). Brooks AFB, TX: Air Force Human Resources Laboratory, Manpower and Personnel Division.

Stanley, J.C. (1971). Reliability. In R.L. Thorndike (Ed.). *Educational Measurement* (2nd ed.). Washington, DC: American Council on Education.

Sternberg, S. (1966). High speed scanning in human memory. *Science, 153,* 652-654.

United States Air Force (1983). *Application procedures for UPT, UPTH and UNT,* Air Force Regulation 51-4. Washington, DC: Department of the Air Force.

Introduction to Back Propagation:
An Artificial Neural Network Paradigm

by

W. A. Sands*
Director, Personnel Systems Research Department
Navy Personnel Research and Development Center
San Diego, California 92152-6800

BACKGROUND

The area of Artificial Neural Networks (ANN) has been receiving rapidly increasing attention in the scientific literature. Three papers have been presented recently to the Military Testing Association. In 1990, Dempsey et al. presented a paper on the use of ANNs for manpower modeling. Last year, the present author gave an introductory paper on ANNs (Sands, 1991) and discussed results of an empirical comparison of ANNs and linear regression analysis (Sands & Wilkins, 1991). While there are numerous types of ANNs, the most widely used paradigm is probably the Back Propagation Network (BPN). This paper provides an introduction to the BPN approach and describes some of the applications of this powerful mathematical modeling technique.

ARTIFICIAL NEURAL NETWORKS

Artificial Neural Networks have been discussed in the literature under a variety of names, including: parallel distributed processing models; connectivist/connectionism models; adaptive systems; self-organizing systems; neurocomputing; and neuromorphic systems (Nelson & Illingworth, 1991, p. 19). Various authors have defined neural networks in different ways. A review of many of these definitions reveals that they have a number of concepts in common. Based upon this review, the following definition is offered. An Artificial Neural Network is a mathematical learning model composed of many simple, distributed processing elements which are organized into layers, communicate through interconnections, and process information in parallel. This definition includes a number of key concepts. An ANN is a mathematical model which is able to learn or adjust, based upon experience with example data. An ANN is based on many, relatively simple distributed processing elements, or nodes, which process information. These nodes are organized into defined layers. There are extensive communication links connecting these nodes. Finally, the nodes within a particular layer operate in parallel, rather than sequentially.

Figure 1 portrays a single processing element. This element is also known as a neuron or node in the ANN literature. Typically, multiple inputs feed into the unit, each weighted by some amount. Usually, these multiple weighted inputs are summed and the resulting value is transformed into a single output signal, using a transfer function. There are several alternative transfer functions, including: linear, threshold, step, sigmoid, and hyperbolic tangent. Generally, a nonlinear transfer function is employed, as a linear transfer function has limited utility. The sigmoid function is a common

* The opinions expressed in this paper are those of the author, are not official, and do not necessarily reflect those of the Navy Department.

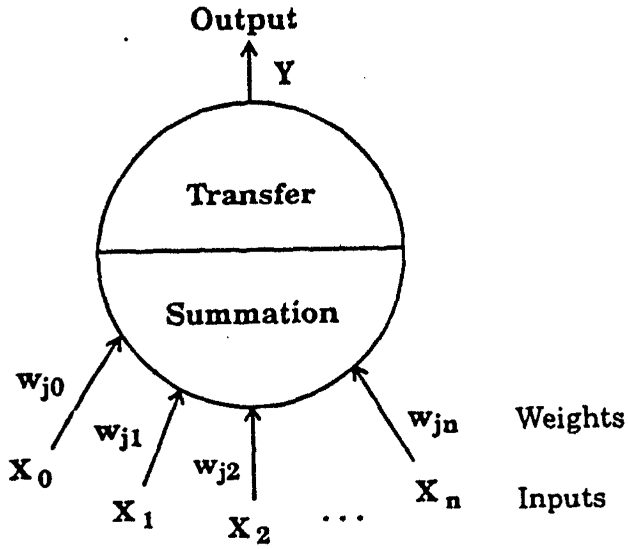


Figure 1. A Single Processing Element.

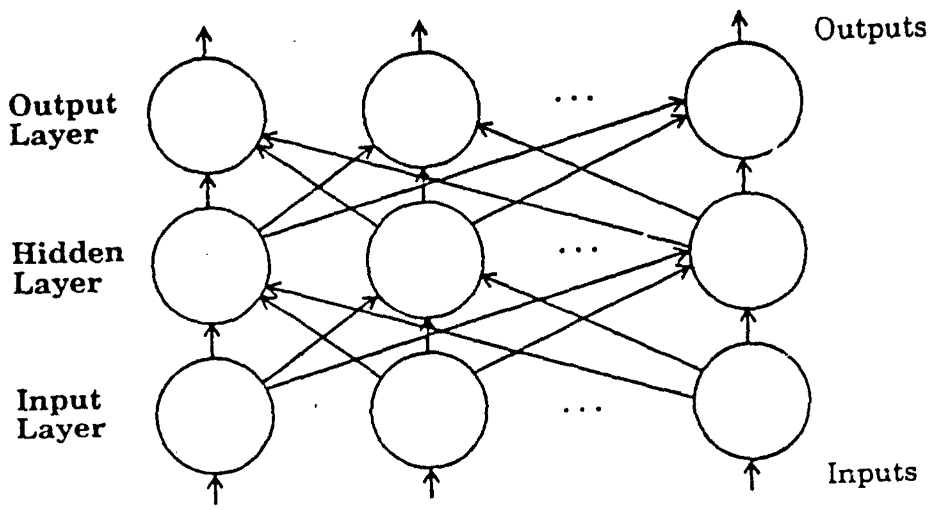


Figure 2. An Artificial Neural Network with multiple processing elements in multiple layers.

choice. This S-shaped function has both a high and a low saturation limit and a proportionality range between these limits.

In general, an ANN involves three types of layers: one input layer, one output layer, and zero, one, or more hidden layers. Each layer can have any number of nodes. As shown in Figure 2, the input nodes obtain information from outside the network and pass it on to the nodes in the hidden layer. These nodes are called "hidden" because they have no direct contact with anything external to the network. After processing takes place in the hidden layer(s), the signals are sent to the output layer nodes, which generate network outputs. If each node in one layer is connected to all nodes in the next higher layer, the network is considered to be fully connected.

The amount of information concerning the correct response that a network is given during training provides a basis for categorizing ANNs. A supervised network is given the correct output for every example case in a training sample. The network output response is compared to the correct response and the difference is considered the error. Connection weights are adjusted to reduce the error on the next iteration through the network. In this way, the network progressively learns to produce the correct response. An unsupervised network receives no information on the correct output response. Rather, these networks monitor their internal performance, looking for regularities, and cluster input examples. Reinforcement learning is a compromise between supervised and unsupervised learning. In this case, only a grade on the quality of the network output is received, not the correct answer.

BACK PROPAGATION NETWORKS

There are many different types of ANNS. Given the wide range of choices, it is important to know which paradigms have a good track record for solving actual problems. At this time, the Back Propagation Network (BPN) appears to be the most popular.

The BPN procedure actually involves two phases. In the forward phase, a vector of inputs is presented to the network at the input layer. Signals are passed to the nodes in the hidden layer, where they are summed to produce a value at each node. These values are then processed through the transfer function (e.g., sigmoid) and the results are passed to the output layer. Then the actual network outputs are compared to the desired, or target output values. The differences between the desired and obtained outputs are considered errors. The backward phase processes the error information back down through the network, changing the connection weights using the Generalized Delta Rule (GDR). The mathematical details of this GDR procedure are available from many publications (e.g., Caudill & Butler, 1992a). In summary, the BPN processing sequence involves two passes through the network. The first phase estimates the error, while the second phase modifies the connection weights to reduce the error.

APPLICATIONS

As Nelson and Illingworth point out (pp. 9, 10), Artificial Neural Networks have been successfully developed in a wide range of application areas, including:

(1) biological - modeling the brain and other biological systems

(2) business - oil probability estimation in geological formations

(3) environmental - weather forecasting

(4) financial - credit risk assessment

(5) manufacturing - assembly line quality inspection

(6) medical - X-ray diagnosis

(7) military - radar signal classification

CONCLUSIONS AND RECOMMENDATIONS

Back Propagation Networks have both strengths and weaknesses. They are efficient to develop, both in terms of time and expense, compared to expert systems. They are good at generalization; i.e., they can extract common features of different cases and ignore the idiosyncratic noise. They are robust and have a high fault tolerance. Finally, they are extremely versatile.

There is a certain amount of trial and error involved in the development of these networks. The training process can become trapped in local minima, and never locate the global error minimum. They are poor at problems requiring precision (e.g., balancing a checkbook) and poor at extrapolation. Finally, it is sometimes difficult to explain the results obtained.

In conclusion, ANN models (especially BPNs) appear to offer considerable promise for personnel research. They should be considered as one tool in a researcher's bag of problem-solving techniques.

REFERENCES AND SUGGESTED READINGS

While not all of the references listed below have been cited in this paper, they are included as suggested reading for those interested in learning more about Artificial Neural Networks.

Caudill, M. (1990). *Neural network primer.* Reprinted from *AI Expert.* San Francisco, CA: Miller Freeman Publications.

Caudill, M. & Butler, C. (1992a). *Understanding neural networks: Computer explorations - Basic networks* (Vol. 1). Cambridge, MA: The MIT Press.

Caudill, M. & Butler, C. (1992b). *Understanding neural networks: Computer explorations - Advanced networks* (Vol. 2). Cambridge, MA: The MIT Press.

Dempsey, J. R., Harris, D. A., & Waters, B. K. (1990). The use of artificial neural networks in military manpower modeling. *Proceedings of the 32nd Annual Conference of the Military Testing Association.* Orange Beach, AL: Naval Education and Training Program Management Support Activity.

Guiver, J. P. & Klimasauskas, C. C. (1991, July/August). Applying neural networks: Part IV - Improving performance. *PC AI.*

Hecht-Nielsen, R. (1990). *Neurocomputing.* Reading, MA: Addison-Wesley Publishing Co.

Lawrence, J. (1992). *Introduction to neural networks and expert systems.* Nevada City, CA: California Scientific Software.

Klimasauskas, C. C. (1991a, January/February). Applying neural networks: Part I - An overview of the series. *PC AI.*

Klimasauskas, C. C. (1991b, March/April). Applying neural networks: Part II - A walk through the application process. *PC AI.*

Klimasauskas, C. C. (1991c, May/June). Applying neural networks: Part III - Training a neural network. *PC AI.*

Klimasauskas, C. C. (1991d, September/October). Applying neural networks: Part V - Integrating a trained network into an application. *PC AI.*

Klimasauskas, C. C. (1991e, November/December). Applying neural networks: Part VI - Special topics. *PC AI.*

Klimasauskas, C. C. (1991f). *Applications in neural computing.* Pittsburgh, PA: NeuralWare, Inc.

McClelland, J. L. and Rumelhart, D. E. (Eds.) (1986). *Parallel distributed processing: Explorations in the microstructure of cognition - Psychological and biological models.* (Vol . 2) Cambridge, MA: The MIT Press.

Nelson, M. C. & Illingworth, W. T. (1991). *A practical guide to neural nets.* Reading, MA: Addison-Wesley Publishing Company, Inc.

Rumelhart, D. E. and McClelland, J. L. (Eds.) (1986). *Parallel distributed processing: Explorations in the microstructure of cognition - Foundations.* (Vol . 1) Cambridge, MA: The MIT press.

Sands, W. A. (1991). Artificial neural networks: A tool for psychologists. *Proceedings of the 33rd Annual Conference of the Military Testing Association.* San Antonio, TX: Armstrong Laboratory Human Resources Directorate and USAF Occupational Measurement Squadron.

Sands, W. A. & Wilkins, C. A. (1991). Artificial neural networks for personnel selection. *Proceedings of the 33rd Annual Conference of the Military Testing Association.* San Antonio, TX: Armstrong Laboratory Human Resources Directorate and USAF Occupational Measurement Squadron.

Sands, W. A. & Wilkins, C. A. (1992, March). A comparison of artificial neural networks and linear regression for dichotomous criterion prediction. In Montague, W. E. (Ed.). *Independent Research and Independent Exploratory Development FY91 Annual Report.* Administrative Publication AP-92-5. San Diego, CA: Navy Personnel Research and Development Center.

Weiss, S. M. & Kulikowski, C. A. (1991). *Computer systems that learn.* San Mateo, CA: Morgan Kaufmann Publishers, Inc.

White, H. (1989). Neural-network learning and statistics. *AI Expert*, Vol. 4, No. 12, p. 52.

Wiggins, V. L., Looper, L. T. & Engquist, S. K. (1991). *Neural networks: A primer*. Technical Paper AL-TP-1991-0011. Brooks Air Force Base, TX: Human Resources Directorate, Manpower and Personnel Division, Armstrong Laboratory.

Wiggins, V. L., Looper, L. T. & Engquist, S. K. (1991). *Neural networks and their application to Air Force personnel modeling.* Technical Report AL-TR-1991-0031. Brooks Air Force Base, TX: Human Resources Directorate, Manpower and Personnel Division, Armstrong Laboratory.

Wilkins, C. A. & Sands, W. A. (In press). *Comparison of a back propagation artificial neural network model with a linear regression model for personnel selection.* Technical Report. San Diego, CA: Navy Personnel Research and Development Center.

The Ultimate Person-Job Match: A Key to Future Worker Productivity

Joe H. Ward, Jr.
Consultant

David. S. Vaughan & Jimmy L. Mitchell
McDonnell Douglas Corporation

Walter E. Driskill
Metrica, Inc.

Hendrick W. Ruck
Human Resources Directorate
Armstrong Laboratory

Abstract

Given current trends of decreasing budgets, reduced manpower, and escalating personnel benefits and training costs, it is imperative that the military services consider human resource policy alternatives that will enhance individual productivity and optimize personal growth and development. In the long run, a highly effective approach would be to focus on improving the match between job incumbents skills and job requirements. This approach begins with the enhancement of the kinds of information available about the jobs and the people as they enter occupations. New technologies are under development to improve the definition of what skills, knowledges, and abilities are required for each occupation; such technology could easily be extended to be job- or position-specific. Using this new technology, a better job requirements data base could be developed as the Base Training System (BTS, formerly the Advanced On-the-job Training System or AOTS) is implemented. Advanced utilization and training models, such as those developed in the Training Decisions Support Technology line of research, could be integrated as well, to help functional managers anticipate future changes in jobs and plan for possible career field transitions. Given recent developments in computers and modeling systems, the merging of all these technologies into an integrated human resources management system becomes both possible and practical. The ultimate system should be able to optimize on multiple functions so as to produce improvements in a variety of outcome variables, such as individual productivity, organizational goal achievement, personal growth and development, and classification structure stability.

Introduction

A great deal has been written and said in recent years about the declining productivity of the American worker, and many solutions have been proposed and tried. Deming, one of the pioneers in this area, emphasized the need to examine organizational processes with the view of maintaining and enhancing quality; he advocated the use of statistical process controls, planning for quality, extensive training of personnel, and the removal of impediments to workers' pride in their work (Deming, 1982). Feigenbaum stressed the need for a new philosophy of management, one which focused on the quality of products and customer satisfaction (Feigenbaum, 1983). Juran (1988) advocated the establishment of task teams to improve product quality. Others have spoken out or written in much the same theme, which has led to a very active and visible movement encouraging productivity enhancement through formal Total Quality Management (TQM) programs.

These and similar quality improvement programs have tended to focus almost exclusively on organizational solutions to the problems of productivity, and have generally ignored the individual worker except as a source of ideas for improvements and as a contributing member of a quality improvement team. Kirsch, Fisher, & Melkunas maintain that TQM is quite different from other approaches in that "it seeks to create an organization of employee teams that are self-managed, self-improving and highly flexible" (1991:221). Yet the key to

individual productivity has to be, in great measure, a focus on the individual worker and helping that worker do a job more efficiently and effectively. Such a focus should be aimed at optimizing the personal growth and development of each individual worker since it is only through such individual growth and development that the quality of the work to be done can be improved. One major thrust in this direction needs to be in better definition of the work to be done; another is to improve the match of individual workers with the available jobs.

Today, people are typically matched to occupations or job categories, rather than to specific jobs. For example, within the Air Force, airmen are classified into Air Force Specialties (occupations) and into six major skill level categories. Airmen are assigned to specific jobs. However, that assignment process makes very little use of specific job requirements or job-oriented personnel information beyond occupation and skill level. Advances in assessing specific job requirements and personnel characteristics, coupled with modern computer technology, make possible a much more sophisticated process for putting individual workers in the specific jobs for which personal productivity and growth can be maximized across an entire organization. The theoretical and mathematical tools for such an optimized person-job matching system have been available for a long time; they are currently being used within the Air Force to assign airmen entering the AF to occupations. With recent advances in job assessment technology, the time is right to consider application of these tools to match individuals to specific jobs. This paper presents an approach for the ultimate, optimized person-job match. First, some key advances in job assessment technology are described. Then an "ultimate" person-job matching approach is presented which makes use of these new job assessment technologies to maximize personal productivity and growth for an entire organization.

Better Defined Job Requirements

Over the last decade, there has been a growing recognition that it is imperative that we develop a better understanding of the tasks that we ask workers to perform. Fleishman and Quaintance have recently noted that:

> There is a need to conceptualize tasks and their characteristics to resolve central problems in the study of human behavior. If we are going to generalize about conditions affecting human performance, it is necessary to consider the properties of tasks as important constructs in psychological research and theory as well as in our conceptions of human work and achievement. Such constructs may help to address many common concerns in basic and applied psychology and to integrate concepts and research in a number of seemingly diverse fields. (1984:1)

Cognitive psychologists have also been drawn to this area. Recently, the American Psychological Association's Science Directorate has been working with the U.S. Department of Labor in a joint project to specify the requirements of occupations to be listed in the next revision of the *Dictionary of Occupational Titles* (DOT). The DOL's Advisory Panel on the Review of the DOT "has concluded that the increasingly cognitive nature of work in the 90's has spurred the need to better understand the mental processes involved in work performance, and to document them as part of the job domain" (APA, 1992:12). This type of approach may, in the long term, provide a much better specification of the kinds of mental skills and abilities required to perform various occupations. In the short term, however, more expeditious methods are needed to gather and process information about the requirements of specific types of work.

Fleishman has recently developed a survey methodology (Fleishman-Job Analysis Survey or F-JAS) to assess the knowledge, skills and abilities (KSA) requirements of jobs, where experienced employees use behaviorally-anchored rating scales to determine how relevant each KSA is to their job (Fleishman and Reilly, 1992). The abilities identified as job requirements using the F-JAS can be directly linked to appropriate ability tests, thus assuring content validity for organizational selection procedures.

In a different approach aimed at generally the same objective, Moon, Driskill, Weissmuller, Strayer, Fisher, & Kirsh (1991) assessed the KSA requirements of Internal Revenue Service jobs by having modules of tasks rated by a panel of subject matter experts (SMEs). Task modules, derived from CODAP statistical clustering based on the co-performance by varying groups of job incumbents, are "excellents units of analysis for determining job requirements" and defining training (Moon, et al., 1992:244). In this study, task module-level KSA linkages provided the framework for assessing content validity of entry-level training and selection tests.

Such linkages ensure job relevance of both tests and training as well as ensuring adequate job coverage.

Task modules have also been used experimentally for gathering ratings of KSA requirements using three different taxomies, for a sample of Air Force occupations (Driskill, Weissmuller, & Dittmar, 1992). These ratings proved to be generally reliable, as assessed by interrater agreement indices, and appeared to realistically differentiate among the specialties. A follow-on project has been planned which will expand this process to a much larger sample of enlisted specialties; if the process continues to be successfully applied, it will eventually lead to a much more systematic approach to establishing Air Force specialty requirements, which could lead to more sophisticated selection and placement testing, as well as more effective training programs.

Advanced occupational analysis software has been developed to assist analysts in the identification and interpretation of both case and task clusters (Phalen, Mitchell, and Hand, 1990). Work continues on refining and testing these programs. The task clustering programs have proved particularly useful, as a way to order the extensive task data base, and make the information more useful to managers and decision makers.

Building A Better Job Requirements Data Base

Once the new technology for collecting job requirements data has been refined, and a systematic approach developed for linking task modules and KSA requirements, we will need to operationalize these methodologies so as to develop a comprehensive job requirements data base for all specialties. Such a data base should include active duty enlisted, civilian, and officer jobs as well as National Guard and Reserve positions. One approach to the development of such a data base is to integrate this requirement with some other new manpower, personnel, and training (MPT) technology being implemented as a way to expedite and systematize the process. One candidate system might be the new base-level system for managing on-the-job training; the Base Training System (BTS; formerly the Advanced On-the-Job Training System or AOTS) which is now being tested by the Human Systems Division (HSD/YARD) for possible Air Force-wide implementation (Blackhurst, et al, 1991).

By integrating the requirement for a new job requirements data base with a system such as BTS which is being implemented, several things can be achieved simultaneously. Since the BTS is designed to provide local supervisors with generic position task lists as a starting point for the development of position-specific OJT programs, the system is oriented toward recognizing major variations in jobs within an occupational field (Air Force specialty). Thus, it will permit the easy identification of those technical supervisors most capable of providing ratings of job requirements for the recognized generic positions, and will therefore facilitate the assessment of variations in such requirements among jobs. When consolidated with data from other locales and units, this permits the systematic evaluation of such job requirements in a way not possible in the past. By obtaining job requirements ratings on task modules, the unit of analysis is shifted from the specialty (AFS) as a whole to the task modules, which can then be reorganized into new jobs as the classification structure changes or as new systems are introduced into the inventory. This provides much greater flexibility to the system, and overcomes some of the major problems inherent in studying job requirements at the specialty level.

Computer-Based Career Field Modeling

Once the more comprehensive data base of job requirements is developed, it can be used for a variety of purposes beyond the development and management of OJT programs. Such data could also be used as another source of information for evaluating possible changes in how an occupation is organized. Advanced utilization and training models, such as those developed in the Training Decisions Support Technology line of research (Mitchell, Vaughan, Knight, Rueter, Fast, Haynes, & Bennett, 1992), can be adapted to make use of such an advanced job requirements data base to help functional managers plan for career field changes. Recent Air Staff initiatives mandate greater responsibility for functional managers in planning and budgeting for all of the training required in their specialties. They are charged with the development of career field training management plans (CFTMPs) to systematize expected transitions based on changes in how career fields are organized, the equipment and systems they operate or maintain, and the training programs needed to make their specialist proficient on the job.

To assist functional managers in making decisions on training and developing CFTMPs to implement their decisions, they may call on representatives of all of the functional areas (by Major Command, base or unit) by convening a Utilization and Training Workshop (U&TW). This type of cooperative effort insures that all major areas of concern can be discussed, and various alternative solutions can be proposed. The Training Decisions

Support Technology is designed to assist functional managers and U&TWs in evaluating alternative career field structures and various configurations of training (Mitchell, et al., 1992). This technology is being extended in an attempt to also assist in the development and drafting of CFTMPs (a report on a pilot project for such a purpose is being presented in another session at this conference).

Given the very rapid development of computer technology over the last few years, it has now become feasible to consider merging many of these technologies into an integrated human resources management system (Mitchell & Driskill, 1986). Most of these various technologies have been developed using tasks (or task modules) as their basic unit of analysis; with that common foundation, integration and interaction is not only feasible but potentially highly effective. The challenge is to develop appropriate interfaces so that data (at some level of abstraction) can be moved easily from one system to another, and to have appropriate checks and balances to insure comparability of results. In this way, the integrity of the basic data can be maintained while various output products are moved from system to system to be used to meet a variety of needs (analysis or restructing of jobs, definition of training requirements, assessment of the impact of proposed changes, etc.).

The Ultimate Person-Job Match

The basic theoretical and mathematical tools for optimized person-job matching have been available for some time and are now being used to make initial assignments for Air Force guaranteed enlistement personnel in the Procurement Management Information System or PROMIS (Ward, 1983; Ward, Haney, Hendrix, & Pina, 1978). Other U.S. uniformed services have applied similar models (Kroeker, 1989). The difficulties have been in assembling the detailed person and job data bases and in the shear magnitude of the computations required for person-job matching. As discussed above, much more sophisticated and detailed data are becoming available concerning job requirements, including data concerning specific tasks and task modules that job incumbents will need to perform. Similarly, much more detailed data are becoming available concerning personnel characteristics, including specific task/task module proficiency data. These person and job data bases, coupled with modern computer technology, make possible the ultimate person-job matching system.

The first step in building the ultimate person-job matching system is to define an objective function. This mathematical function quantifies the relative value or utility of placing a given person into a specific job. Its independent variables will be person and job characteristics that predict or lead to the relative value of an assignment. A given person-job match may have different values, when viewed from different perspectives. One type of value relates to a person's productivity in a given job. This value would probably be maximized by improving the match between a person's current task proficiencies and a job's required tasks. Another type of value might relate to a person's preference for a particular job or assignment. This might involve such variables as geographic location, which are unrelated to job task performance requirements. A third type of value might relate to expansion of personal experience and skill, to prepare for future jobs.

The objective function reflects trades among these various (possibly conflicting) types of values associated with a person-job match. While this trade can be mathematically reflected in a variety of ways, perhaps the simpliest involves a weighted linear function of measures for the different values. Weights may also be required within a single value measure. For example, some tasks in a job may be more important for overall job productivity than other tasks. These tasks should receive more weight in scoring the match between a person's task proficiency and a job's task requirements. Similarly, achieving a good match may be more important for some jobs than for other jobs; these jobs can be weighted more heavily in the overall value measure.

A key aspect in the success of a person-job matching system involves determining a set of weights that is acceptable to all parties. Such weights are subjective in nature and reflect a compromise between several different points of view. They can be determined using policy capturing/policy specifying methods (Ward, 1977).

The next step in building the ultimate person-job matching system involves defining constraints that an acceptable organization-wide set of person-job matches must meet. One constraint, for example, is that exactly one person be placed in each job. Constraints may also be related to personnel policies and processes. For example, each person who is currently in an overseas assignment might be required to be matched to a CONUS assignment. Exceptions to constraints can be made as appropriate. For example, if an individual who is currently overseas wanted to stay overseas, the CONUS assignment constraint could be voided for that individual.

The information on individuals currently maintained in the Personnel Data System has been in very general terms (e.g., only the last three training courses taken, only job title and duty AFSC for assignment history, etc.). The development of the BTS will help solve the latter problem in that an individual training record (ITR) will

be maintained which documents the jobs and individual was assigned and the training received OJT to prepare the individual to perform that job. The individual's record will include very specific information in terms of proficiency achieved in accomplishing specific tasks or groups of tasks. The ITR will provide a wealth of information for the individual supervisor, as a basis for specifying the required OJT program. We will need, however, some system to periodically pull ITR data into a central AFS-specific data base, if such information is to be used for AFS modeling or person-job match application.

The needs of the Air Force, in terms of priorities (or criticality) for manning some positions, can be weighted into the equation, as can the preferences of the individual (in terms of an ordered list of preferred assignments). Given that the right kinds of information are available, and some priorities established (in terms of which needs or requirements should be more heavily weighted), there is no reason that multiple functions cannot be used in an optimization algorithm. The mathematics required for such a system have been available for many years; the data bases needed, however, are just now being formulated. Such data bases are complex in terms of many variables and many data points, but are generally straightfoward once the requirements have been defined.

With data bases of person and job data and with appropriate objective and constraint functions defined, computer algorithms can be applied to match people to jobs in order to optimze the objective function and meet the constraints. In large occupations, this computational problem may be substantial. Another implementation issue concerns the time window within which people and jobs are collected for matching. If this time window is too narrow, few people and jobs will be collected; this will limit the degree of match that can be achieved.

The other main problem in the past has been the issue of computing power; the capability to manipulate large data bases in a variety of ways to achieve multiple objectives. Recent advances in computers and minaturization have solved this problem. Today, micro-computers are available at reasonable cost which can accomplish large data base manipulations which use to consume days and weeks of computer time on large mainframe systems. Even some Personal Computers (PCs) now have the computing power which used to reside only on mainframes. Software systems, such as typical regression programs or CODAP, are rapidly being modified so that they can be used on several different size systems, from the largest mainframes to the top-of-the-line PCs. Given this kind of capability, computing power is no longer an excuse for delaying implementation of advanced person-job match technology.

Conclusions

As these new support technologies (e.g., BTS, TDS, CFTMPs, U&TWs, etc.) are operationalized, the new job requirements and training data bases will rapidly become available for research and development of needed interfaces and creation of an integrating system. This opportunity is one which should not be missed, for it will aid functional and MPT planners and decision makers in creating better job specifications and better training programs for the reduced work force. If we apply the basic concepts of person-job match technology in future Air Force and DoD MPT operations, the result will be a highly trained, proficient work force; people who know what and where they are, and where they are going. In its penultimate application, person-job match technology (philosophy) will lead to substantially enhanced individual productivity and professionalism. This is, after all, the basic principle inherent in the PJM system as well as the whole TQM movement. A realistic implementation of PJM is needed if we are ever to achieve full TQM success.

References Cited

American Psychological Association (1992, September/October). Joint APA, DOL Report will investigate applications of cognitive task analysis. *Psychological Science Agenda* 5(5):12. Washington DC: APA, Science Directorate.

Blackhurst, J., Mitchell, J.L., Ruck, H.W. & Driskill, W.E. (1991). Integrating R&D results: A model for managing implementation of new MPT processes. *Proceedings of the 33rd Annual Conference of the Military Testing Association* (pp. 348-353). San Antonio, TX: Armstrong Laboratory Human Resources Directorate and the USAF Occupational Measurement Squadron.

Deming, W. E. (1982). *Quality, productivity, and competitive position*. Cambridge, MA: Massachusetts Institute of Technology Press.

Driskill, W.F., Weissmuller, J.J., & Dittmar, M.J. (1992). Enlisted job classification technology: Initial development. Presentation to the 34th Annual Conference of the Military Testing Association Conference, San Diego, CA.

Feigenbaum, Armand (1983). *Total quality control*. Cincinnati, OH: Association for Quality and Participation.

Fleishman, E.W., & Quaintance, M.K. (1984). *Taxonomies of human performance: The description of human tasks*. Orlando, FL: Academic Press, Inc.

Fleishman, E.W., & Reilly, M.E. (1992). *Handbook of human abilities*. Palo Alto, CA: Consulting Psychologists Press, Inc.

Juran, J.M. (1988). *On planning for quality*. Cambridge, MA: Productivity Press.

Kirsch, M., Fisher, G., & Melkunas, C. (1991). Using CODAP as a quality improvement tool. *Proceedings of the 33rd Annual Conference of the Military Testing Association* (pp. 221-225). San Antonio, TX: Armstrong Laboratory Human Resources Directorate and the USAF Occupational Measurement Squadron.

Kroeker, Leonard (1989). Personnel\Classification Assignment Models. In (Wiskoff, M..F. & Rampton, G.M., Editors) *Military Personnel Measurement: Testing, Assignment, Evaluation* (pp. 41-73). New York: Praeger.

Mitchell, J.L., & Driskill, W. E. (1986, August 25). Optimizing integrated personnel system training decisions and development. Presentation in the symposium, State-of-the-Art Applications of Job Analysis: Integrated Personnel Systems, American Psychological Association convention, Washington, D.C,

Mitchell, J.L., Vaughan, D.S., Knight, J.R., Rueter, F.H., Fast J., Haynes, W.R., & Bennett, W.R. (1992). *Training decisions technology analysis* (AL-TP-1992-0026). Brooks AFB, TX: Technical Training Research Division, Armstrong Laboratory, Human Resources Directorate.

Moon, R.A.J., Driskill, W.E., Weissmuller, J.J., Strayer, S.J., Fisher, G.P., & Kirsh, M. (1991). Using task co-performance modules to define job requirements. *Proceedings of the 33rd Annual Conference of the Military Testing Association* (pp. 243-252). San Antonio, TX: Armstrong Laboratory Human Resources Directorate and the USAF Occupational Measurement Squadron.

Phalen, W.J., Mitchell, J.L., & Hand, D.K. (1990). ASCII CODAP: Progress report on applications of advanced occupational analysis software. *Proceedings of the 32nd Annual Conference of the Military Testing Association* (pp. 82-87). Orange Beach, AL: Naval Education and Training Program Management Support Activity.

Ward, J.H., Jr. (1977, August). *Creating mathematical models of judgement processes: From policy-capturing to policy-specifying* (AFHRL-TR-77-47, AD-1048 983). Brooks AFB, TX: Occupational and Manpower Research Division, Air Force Human Resources Laboratory [Also *Journal of Experimental Education*, 48(1):60-84, 1979].

Ward, J. (1983, April). Strategies for capitalizing on individual differences in military personnel systems. In R.C. Sorenson (Ed), *Human Individual Differences in Military Systems* (NPRDC SR 83-30). San Diego, CA: Navy Personnel Research and Development Center.

Ward, J.H., Haney, D.L., Hendrix, W.H., & Pina, M. (1978, July). *Assignment procedures in the Air Force procurement management information system* (AFHRL-TR-78-30, AD-A056 531). Brooks AFB, TX: Occupation and Manpower Research Division, Air Force Human Resources Laboratory. [Also *Journal of Experimental Education*, 47,(2): 149-155., 1978].

An Integrated Training Systems Model for Modular Control Equipment

Elizabeth G. Gibson
Training Technology Division
Mei Technology Corporation

Introduction

Assessing the training requirements of pre-production equipment involves specific training type and timing issues. This paper illustrates the methods used and findings associated with an evaluation of Computer-Based Training (CBT) for the U.S. Air Force (USAF) AN/TYQ-23(V)2 Modular Control Equipment (MCE). This includes the identification of skills and knowledge suitable for CBT, definition of CBT system requirements, selection and justification of CBT system development options, and categorization of learning components representing a practical and theoretical approach which can be applied to modeling any new training system.

Background

The USAF is replacing aging tactical air operations 407L control facilities with updated modular equipment manufactured by Litton Data Systems. The MCE is a small, compact air traffic and air defense control center housed in an 8'x 8'x 20'shelter. It is the major component of the Tactical Air Operations Module (TAOM), which links air weapons controllers, pilots, and ground communications personnel. MCE operators perform surveillance, identification, automatic target acquisition, tracking, and threat evaluation and share information with other systems via tactical data links. Each shelter contains four operator console units and their associated radar processors, communications equipment, and computers, and has various deployment options. Modules will normally be deployed to forward positions.

Four identical operator console units (OCUs) in each shelter provide the main human-machine interface in each MCE. Operators can perform different functions (e.g., battle management, weapons control, surveillance) tailoring their positions by selecting different switch menus and specific task-oriented displays. Each OCU includes a 25" CRT radar graphics display, a system access subunit, and two auxiliary display subunit screens with a reconfigurable menu panel. Both displays are touch screen activated. Other OCU components include a fixed function control panel, a system status display unit, and a voice communications access unit.

This unusual equipment configuration offers enhanced system capabilities and challenges to human information processing, performance capabilities, and training. The OCU displays a wide variety of data in different formats which are updated at frequent intervals. Data processing requires constant shifts in behavior, including cognitive activities (e.g., monitoring, diagnosis, decision-making), and perceptual motor switch action-based activities.

Acquiring the skills and knowledge to effectively operate and interact with this complex equipment, computer software, and the environment in which the system operates, can be

facilitated by the individual approach to learning and evaluation provided by CBT technology. The use of CBT is increasing throughout the USAF. The term "CBT" is used here to refer to all training technologies which utilize a personal computer (PC) as a centerpiece, including interactive video, Digital Video Interactive (DVI), and PC-based simulations. It also encompasses other technologies such as Computer-Assisted Instruction (CAI), Computer-Based Instruction (CBI), and Computer-Managed Instruction (CMI). This research focuses on the application of CBT technology to potentially increase MCE training performance, efficiency and cost effectiveness.

Methods

The principal research method employed was to determine the MCE training requirements which were suitable for CBT, then to describe alternative hardware/software configuration to meet these requirements. Learning objectives were evaluated to determine their appropriateness for CBT, to determine skill/knowledge commonalities among objectives, and to rank order them according to their priority ratings. Once analyzed, objectives were grouped into learning modules for training and prioritized for development.

Evaluation of the MCE training system included: 1) Identification of skills and knowledge needed to support the system; 2) Prioritization of Initial Skills Qualification Training (ISQT) and On-the-Job training objectives for the application of CBT technology; 3) Recommendation of CBT learning modules to be developed; 4) Emulation of MCE consoles for CBT to guide students through switch actions (e.g., complex sequence of touchscreen panel presses onto a dual-screen console with trackball and keyboard input devices); 5) Rank ordering of various CBT hardware and software configurations as to their capability to address the MCE training requirements, and; 6) Selection and justification of an appropriate training system for MCE. Media selection criteria applied to these objectives included conditions, standards, potential changes, type of training, instructional setting, learning environment, equipment specifications, requirement for standardization, delivery method, and level of student interaction (Walsh, Yee and Young 1991).

Operational documentation and procedural manuals on MCE were reviewed to gather data on the background, configuration, jobs/tasks, objectives, and learning activities involving the new equipment. The Occupational Survey Report for the 17XX Air Weapons Controller career specialty field was used to verify a baseline of MCE objectives. Lesson plans from the ISQT "factory course" were also evaluated to confirm what tasks were required and how tasks were performed. Several site visits were made to observe, compare and contrast operators using both the 407L system and MCE. Subject matter experts (SMEs) from the 607 Tactical Control Training Command (TCTS), Luke AFB, verified, revised and assigned ratings to the list of MCE objectives and their associated training characteristics, including task difficulty, frequency, and criticality.

Findings

MCE learning objectives, based on knowledge, skill, and performance-based behaviors, were categorized based on whether or not they could be taught and/or tested using various CBT modes, such as tutorial, drill and practice, or simulation. Some skill and performance objectives which require performing procedures on the actual equipment, such as setting up a radar unit or

timing the performance of a series of switch actions on a console may not be appropriate for CBT, even though many associated knowledge-based objectives are capable of being trained and tested under the simulated conditions of CBT.

After media selection, tasks were prioritized and assigned weights according to criteria described in Table 1. Switch actions, ISQT, and CBT simulation capability were assigned the most weight; the maximum weight for each objective was 11 points.

TABLE 1. MCE Objectives Weighting Factors

Factor/Weight	Rationale
Switch Action/3	Most critical factor in training MCE operators; a weighted factor of 3 was assigned to those objectives which involved the performance of switch actions because of the high degree of importance which is associated with switch actions and the proper operation of the MCE equipment.
ISQT/2	High degree of importance associated with ISQT training due to the fact that the majority of new MCE operators will be required to attend ISQT training; of necessity Continuation Training receives lower ranking.
Simulation/2	Those objectives which are capable of being simulated via CBT were identified and given a weight of 2; objectives which can be simulated will help alleviate pressure and bottlenecks in training on the actual equipment.
Common Task/1	Objectives for tasks which are performed by more than one position within the operating module were assigned a weight of 1.
Difficulty/1	Objectives for tasks which are performed during emergency situations, i.e., under pressure or tasks which have long or intricate procedures were assigned a weight of 1; all other objectives received no weight for this factor.
Frequency/1	Objectives for tasks which are performed on the basis of more than once/session were assigned a weight of 1; all other objectives received no weight for this factor.
Criticality/1	Objectives for those tasks which are critical to mission accomplishment were given a weight of 1; all other objectives received no weight for this factor.

Table 2 shows the distribution of ratings and percentage of MCE objectives receiving each ranking based on the criteria listed in Table 1. Learning objectives were grouped into several layers based on these rankings, which reflected their importance for inclusion in the learning modules. Seven learning objectives (.8%) were given the highest rating of "10." An

example is the objective "Respond to specified ALERTS," which was appropriate for ISQT, a switch action, part of simulation, and which was a common, difficult, and frequent task. The most common rating, "2" was assigned to 288 (32.2%) of the learning objectives, for example, "Select the TRACKING MODES" was identified as being part of the ISQT curriculum.

TABLE 2. Distribution of MCE Objectives Ratings

Rating	Number of Objectives	Percentage
10	7	.8%
9	47	5.2%
8	37	4.1%
7	102	11.4%
6	43	4.8%
5	11	1.3%
4	268	29.9%
3	92	10.3%
2	288	32.2%
Total:	895	100%

After learning objectives were ranked individually, the highest priority Terminal Learning Objectives (TLOs) (i.e., ranked 10) were grouped with their associated Enabling Learning Objectives (ELOs) to form learning modules. Some skills and knowledge need to be learned before other higher level behaviors. For example, a weapons controller must learn the definition of HOOK DATA READOUT (HDRO) before learning how to interpret the various scenarios. Grouping learning objectives and learning events into coherent modules enhances training. The top eight learning modules and their estimated training time are listed in Table 3.

TABLE 3. List of Highest Priority CBT Learning Modules

Learning Module	Learning Module Name	Training Time
1	Hook Data Readout	3 hr 15 min
2	System Error Messages	1 hr 30 min
3	Alerts	5 hr
4	Operator Console Unit Set Up	11 hr 15 min
5	Printer Unit	1 hr 45 min
6	Voice Communication Unit	1 hr 15 min
7	Symbology Display	3 hr 15 min
8	Common User Fixed Function Switches	26 hr

A sample learning module for "HOOK DATA READOUT," including brief descriptions of the seven objectives, priority rating, and CBT format, is shown in Table 4. This module contains a mix of training types including knowledge and skill-based learning objectives. Different training formats provide varying levels of opportunities for interaction, feedback, and fidelity to the operational environment. CBT simulation was selected for high priority switch action TLOs. Supporting knowledge and skills ELOs were appropriate for tutorials, and functionally related ELOs were assigned to practice sessions. As a general rule, objectives best trained using a CBT simulation were allotted one hour, a part-task training or practice session for an enabling objective, 30 minutes, and for most other enabling or academic objectives, called "tutorial," 15 minutes each. The total time estimated to present this learning module is 3 hours and 15 minutes or about four traditional class periods. If students cannot master the objective, they can receive additional practice by running the CBT simulation or specific tutorial again.

TABLE 4. HOOK DATA READOUT Learning Module

Objective Description	Priority Rating	Training Type	Training Time
Interpret HOOK DATA READOUTS (HDROS)	10	Simulation	1 hr
Match the proper definition to the term "HDRO"	4	Tutorial	15 min
Match the HDRO areas displayed on the Auxiliary Display Subunit-1 (ADS-1) to the proper definition	4	Tutorial	15 min
Select the HDRO amplifying data types	4	Tutorial	15 min
Read the Common Data area of the HDRO	4	Practice	30 min
Read the Amplifying Data area of the HDRO	4	Practice	30 min
Read the Special Data area of the HDRO	4	Practice	30 min
		Subtotal	3 hr 15 min

The second major task for MCE training was to determine hardware and software configurations capable of presenting the objectives selected for CBT based on MCE training requirements. The most important consideration was the relationship between the training mode and the fully operational mode and its implications for learning transfer. Two touchscreen monitors are primary components of the operational environment; touching the screens using one

hand on each screen is the primary mode of operator interaction. Other significant factors considered were if the system could use existing equipment and could the system used for CBT also be used to simulate MCE operations.

CBT authoring systems (for creating and running the programs) were assessed based on their user interface capabilities for ease of courseware development and maintenance, ability to display different information on two screens, and student management functions. Compatibility with potential simulator devices for the MCE was also a factor. Although several configurations were acceptable, the most practical recommendation was a combination of IBM/PC-compatible machines already on-board and a DOS-based authoring system.

Discussion

CBT can be a viable, efficient, and cost-effective alternative to traditional training systems and high cost simulators when learning objectives are systematically evaluated and grouped into comprehensive learning modules. In general, CBT is an excellent training medium for knowledge and skill-based objectives because: 1) It can provide high-level interaction, graphics and video presentation, self-paced study and remediation, and; 2) It can simulate most equipment and teach students where different components are located and how they function. There are other variables, besides the types of learning objectives, which are important to consider when developing a new training system. For example, some objectives may require special training settings, environmental conditions, or operational equipment which may exclude the CBT medium; others depend on specific behavioral measurements, such as assessment of mechanical or psychomotor skills, which may require another delivery medium to facilitate task transfer. Different training types (i.e., formal school, OJT, field) may require different physical settings, levels of student interaction, or learning environments which may or may not be appropriate for CBT (Walsh and Gibson 1991).

This paper describes a model, using the MCE training system as an example, which incorporates a task analysis and Instructional Systems Design (ISD) approach, including dynamic training characteristics and other pertinent data from users. Relative importance and scaling variables permit users to make informed training decisions based on various cost-benefits of development effort and training fidelity. This is a practical approach for comparing and combining training system components to create the most appropriate configuration for achieving the learning objectives with applications in a broad range of training contexts. Future work will involve designing, developing, and validating lessons and the summative evaluation of results.

References

Walsh, W.J. and Gibson, E.G. (1991). Modular Control Equipment (MCE), Computer-Based Training (CBT) Systems Trade Study, Final Report. San Antonio, Texas: Mei Technology Corporation. Prepared for USAF, Headquarters, Electronic Systems Division (ESD/TCM), Hanscom AFB, Massachusetts, under Contract No. F33615-88-C-003/0017.

Walsh, W.J., Yee, P.J., and Young, S.A. (1991). Characterization of Air Force Training and Computer-based Training Systems. USAF Technical Report #AL-TP-1991-0048, Brooks Air Force Base, Texas.

AN INITIAL EVALUATION OF THE USE OF CAPTIONED TELEVISION TO IMPROVE THE VOCABULARY AND READING COMPREHENSION OF NAVY SAILORS

Ray Griffin and Jeanie Dumestre

NAVAL EDUCATION AND TRAINING PROGRAM MANAGEMENT SUPPORT ACTIVITY
PENSACOLA FLORIDA

INTRODUCTION AND PROBLEM

The Navy needs reliable methods to improve basic academic skills of a portion of its enlisted workforce. Today's sailors must operate and maintain some of the most sophisticated equipment in existence. Yet in 1990 and 1991 about one in four recruits entered the Navy with reading skills below the ninth grade level. While the down-sizing of the military will reduce accessions over the next few years, there will still be many persons in need of improved basic skills.

Captioned TV is the process of presenting the audio track of programs visually, on the TV screen. Closed-captioning was developed to enable deaf and hearing impaired persons to understand TV. Recently, it has been used to try to improve the vocabulary and reading skills of functionally illiterate adults. When used with adults, both the captions and the audio portion of the TV program are presented together initially. Then the volume is reduced. Viewers must read the captions to follow the program. In this way the technique builds on the interest and motivational aspects of television to prompt viewers to read.

At present, a wide variety of television programs are captioned, including news, documentaries, dramas, movies, sports events, and advertisements. As a result, educators may choose from a variety of material for use with language learners of different ages and interests.

REVIEW OF THE LITERATURE

Most of the literature has reported on the use of captions to make TV understandable for the deaf or hearing impaired. Recently, captioned TV has been used to attempt to improve the reading and vocabulary skills of hearing subjects. Much of this literature documents the use of captioned TV to teach reading skills to persons whose native language is not English. These studies are referred to in the literature as "ESL," meaning English as a second language. Results, (Center for Applied Linguistics 1989, Layton 1991, Mehler 1988, Spanos and Smith 1990), show that captioned TV is a viable method for improving English language understanding and speaking of ESL children and adults.

There are only a few formal studies concerned with improving

reading and vocabulary skills of persons who are neither hearing impaired nor ESL in the literature. Jensema, Koskinen and Wilson (1984) tested the use of captioned TV in a University of Maryland Reading Center. The evaluation was conducted after an earlier study resulted in enthusiastic support for captioned TV by ESL and non-ESL remedial reading students. Thirty-five students in grades two through six, and ten teachers, participated in the study. Seventy-five percent of the subjects were remedial readers with an average third grade reading level. Comments about the captioned TV technique were collected from teachers and students using a questionnaire.

Teacher responses were positive averaging 4.2 on a five point Likert scale. Student responses showed that all liked watching captioned TV. Eighty-nine percent of the students said that the technique helped them learn more words. An identical percentage of students said they would like to learn using captioned TV lessons in school. A smaller proportion, 77 percent, said they understood better when watching captioned TV.

Koskinen, Wilson, Gambrell, and Jensema (1986) conducted a follow-on study of 77 learning disabled students in four Maryland public schools. The students were fourth graders who ranged in age from 9 to 13 and were reading at least two years below their grade level. The students were assigned to four groups: closed captioning with sound, closed captioning without sound, television without closed-captioning, and written text only. Lessons were based on the children's TV program 3-2-1 Contact. Student performance was evaluated using word recognition, cloze, silent comprehension, and oral reading tests. There were no significant word recognition differences among the four treatment groups. A subsequent evaluation was made for students reaching a 90 percent criterion for word recognition. The subjects of the captioned TV with sound group, performed better than the text only, or captioned TV without sound, groups.

Bean (1989) used closed captioned TV to teach reading to 24 adults attending a literacy program. Reading achievement scores were used to place subjects into three instructional groups of equal reading ability. The groups were: closed captioned TV with vocabulary instruction, script with vocabulary instruction and closed captioned TV without instruction. Segments from the TV program 3-2-1 Contact were used in the study. All student groups showed improved performance on post-treatment word recognition tests. And, there were no differences between the instructional methods. The authors noted that subjects of the captioned TV without additional vocabulary instruction group performed as well as the captioned TV with instruction group. In addition, 100 percent of the students receiving the captioned TV treatment said that they found the medium enjoyable. The author concluded that captioned TV has potential for aiding sight vocabulary as well as being an enjoyable and motivational method of learning.

In conclusion, results on which to judge the value of

captioned TV to improve reading comprehension and vocabulary skills of adults are few. However, individuals receiving closed captioned TV instruction are enthusiastic about it.

METHOD

Individuals in this study were first term enlisted sailors on board the USS LEXINGTON in Pensacola, Florida. Potential subjects were identified on the basis of their Armed Services Vocational Aptitude Battery "Verbal" score as at or below the ninth grade reading level. These selected sailors were presented the history and use of closed captioned TV, and told the objectives of the study. They were asked to volunteer to become a subject. Incentives for participation were a 72 hour pass for experimental subjects and additional sleep time for controls. The resulting volunteer subjects were randomly selected to receive either the control or experimental treatment.

Each volunteer was tested using the Tests of Adult Basic Education (TABE) level A Form 5. This test served to identify the baseline vocabulary and reading skill of the subjects. Initial skill levels would be compared with those obtained after participation in the captioned TV experiment.

The experimental subjects participated in a series of 36 captioned TV sessions aboard the USS LEXINGTON between February 11 and April 19, 1991. The sessions consisted of viewing and reading the audio portions of the TV programs Highway to Heaven or the Oprah Winfrey Show. The sessions were presented each afternoon between 15-1600 hours (3-4 O'clock). Each session lasted about one hour, including commercials. Many of the commercials were also captioned. The session procedure consisted of turning on the television program and after a few seconds, turning down the volume so it was inaudible. This procedure encouraged subjects to read the visual presentation of the normally audible soundtrack. At least one research monitor was present at each experimental session to observe subject participation.

An equally difficult form of the TABE (Level A Form 6), was given to control and experimental subjects on the fourth day after the end of the captioned TV sessions. Test performance was analyzed to determine reading and vocabulary gains (if any) resulting from the experimental treatment.

RESULTS

The TABE Level A Form 6 was administered to the 29 experimental subjects to determine any vocabulary or reading comprehension gains resulting from the experimental treatment. The test was also given to the 16 control subjects to assess any performance changes over the period of the evaluation.

Most of the experimental and control subjects were male--97

percent in the experimental,and 75 percent in the control groups. There were five females, four in the control and one in the experimental groups. Also, most of the control (56 percent) and experimental (79 percent) subjects were minorities.

A comparison of pre-and-post vocabulary and reading comprehension test performance using t tests for independent and dependent (correlated) means showed no significant differences between the experimental and control groups in pre-or-post test performance.

These results suggest that no vocabulary or reading comprehension gains resulted from the captioned TV treatment. However, the subjects attended a wide range (from 1 to 32) of captioned TV sessions. The attendance average was seventeen sessions. Attendance at only one session would not be enough to result in improved vocabulary or reading comprehension skills. However, the needed number of sessions to effect change was unknown. As a result, test means were compared for subjects who attended 11 or more, and 20 or more sessions, respectively. The resulting means, t and p values are in Tables 1 and 2.

The results for the 19 sailors attending 11 or more sessions (see Table 1) reveal a slight but significant increase in the vocabulary performance of these experimental subjects. Their scores increased from 18.32 on the pre-test to 19.68 on the post-test: $t = 2.06$, $p = .05$. Controls performed slightly higher on the pretest than did the experimental subjects but did not experience a significant increase in vocabulary test performance. Control subject pre-and-post vocabulary scores were 19.44 and 19.94 respectively.

The vocabulary scores for the 12 experimental subjects attending 20 or more sessions (see Table 2), increased from 17.92 on the pre-test to 20.08 on the post-test: $t = 2.54$, $p = < .05$. Neither control nor experimental subjects made any significant gains in reading comprehension test scores.

CONCLUSIONS

The 19 subjects attending 11 or more captioned TV sessions improved their vocabulary performance based on a pre-and-post test: $t = 2.06$, $p = .05$. In addition, the 12 subjects attending 20 or more sessions experienced even greater vocabulary test gains: $t = 2.54$, $p = < .05$. There was no similar improvement in reading comprehension based on increased session attendance. Control subject pre-and-post test performance showed no significant change over the period of the experimental treatment. These initial results suggest that given enough treatment, captioned TV may increase the vocabulary test performance of Navy sailors.

TABLE 1. Experimental subjects attending 11 or more captioned TV sessions: Descriptive Statistics, t and p values (two-tailed), 19 experimental and 16 control subjects.

INDEPENDENT MEAN COMPARISONS - BETWEEN GROUPS				
Test Measure	Experimental Mean/Std.Dev	Control Mean/Std.Dev	t	p
Vocabulary pre-test	18.32/2.98	19.44/4.73	0.85	.40
Vocabulary post-test	19.68/3.77	19.94/4.37	0.18	.85
Reading Pre-test	26.32/3.87	27.94/4.67	1.12	.27
Reading Post-test	26.68/4.92	26.88/6.67	0.10	.92

REPEATED MEASURE MEAN COMPARISONS - WITHIN GROUPS				
Test Measure	Pre-test Mean/Std.Dev	Post-test Mean/Std.Dev	t	p
Experimental Subjects				
Vocabulary	18.32/2.98	19.68/3.77	2.06	.05*
Reading	26.32/3.87	26.68/4.92	0.30	.77
Control Subjects				
Vocabulary	19.44/4.73	19.94/4.37	0.55	.59
Reading	27.94/4.67	26.88/6.67	0.86	.40

TABLE 2. Experimental subjects attending 20 or more captioned TV sessions: Descriptive Statistics, t and p values (two-tailed), 12 experimental and 16 control subjects.

INDEPENDENT MEAN COMPARISONS - BETWEEN GROUPS				
Test Measure	Experimental Mean/Std.Dev	Control Mean/Std.Dev	t	p
Vocabulary pre-test	17.91/3.26	19.44/4.73	0.95	.35
Vocabulary post-test	20.08/4.01	19.94/4.37	0.09	.93
Reading Comp Pre-test	26.75/3.41	27.94/4.67	0.74	.46
Reading Comp Post-test	27.75/4.79	26.88/6.67	0.38	.70

REPEATED MEASURE MEAN COMPARISONS - WITHIN GROUPS				
Test Measure	Pre-test Mean/Std.Dev	Post-test Mean/Std.Dev	t	p
Experimental Subjects				
Vocabulary	17.91/3.26	20.08/4.01	2.54	.03*
Reading Comprehension	26.75/3.41	27.75/4.79	0.68	.51
Control Subjects				
Vocabulary	19.44/4.73	19.94/4.37	0.55	.59
Reading Comprehension	27.94/4.67	26.88/6.67	0.86	.40

REFERENCES

Bean, R. M. (1989). Using Closed Captioned Television to teach reading to adults. Reading Research and Instruction, 28 (4), 27-37.

Center for Applied Linguistics. (1989). Evaluating the benefits of Closed-Captioned TV programming as instructional material for ESL students. (Final Report). Washington D. C.

Jensema, C., Koskinen, P., and Wilson, R. (1984). Teaching reading to hearing children via Captioned Television. Computers, Reading, and Language Arts, Summer/Fall.

Koskinen, P. S., Wilson, R. M., Gambrell, L. B., and Jensema, C. J. (1986). Closed-Captioned Television: A new technology for enhancing reading skills of learning disabled students. ERS Spectrum, IV, (2), 9-13.

Layton, K. (1991). Closed-Captioned Television: A viable technology for the reading teacher. Reading Teacher, 44, (8), 598-599.

Mehler, A. (1988). The potential of Captioned Television for adult learners. Working Papers of Planning an Development Research--Working Paper 88-3. TV Ontario, Toronto.

Spanos, G. and Smith J. (1990). Closed Captioned Television for adult LEP literacy learners. Eric Digest, National Clearinghouse on Literacy Education. Washington D. C.

STUDENTS' RECEPTION AND DISCOVERY METHODS AFFECTED BY THEIR INFORMATION PROCESSING BEHAVIOR

by

Waymond Rodgers
Associate Professor
Graduate School of Management
University of California, Riverside
Irvine, CA 92521

October 1992

ABSTRACT

The objective of this study is to relate students' reception and discovery processes with what and how financial information is processed before they make decision choices. The issue here is to measure the effects of different types of information on students' cognitive processes. Students may encode information at very different levels of processing. Two learning theories of reception and discovery were used in this study to depict students' first and second stage processes, respectively. To determine if certain students process information at different levels, an intolerance-of-ambiguity test was used to divide the students' processes into two groups. Graduate students were presented with ten independent cases of information and were told to make financial decisions. The results indicated that the ambiguity intolerant students outperformed the ambiguity tolerant students. These results implied that the ambiguity intolerant students relied more on reception learning methods than did the ambiguity tolerant students. That is, ambiguity intolerant students appear to have internalized or incorporated the information better than the ambiguity tolerant types. The results of this study present another way to analyze students' behavior by examining perceptual and judgmental processes in a covariance structural model.

Figure 1

Decision Making Process

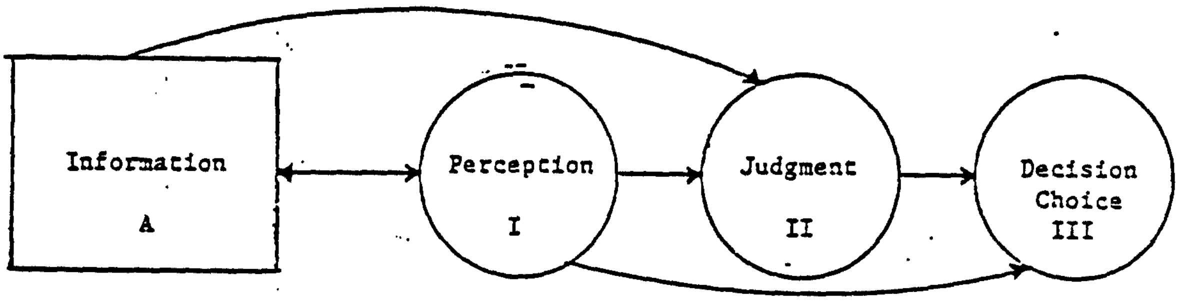


Figure 2

Reception and Discovery Learning Methods Model

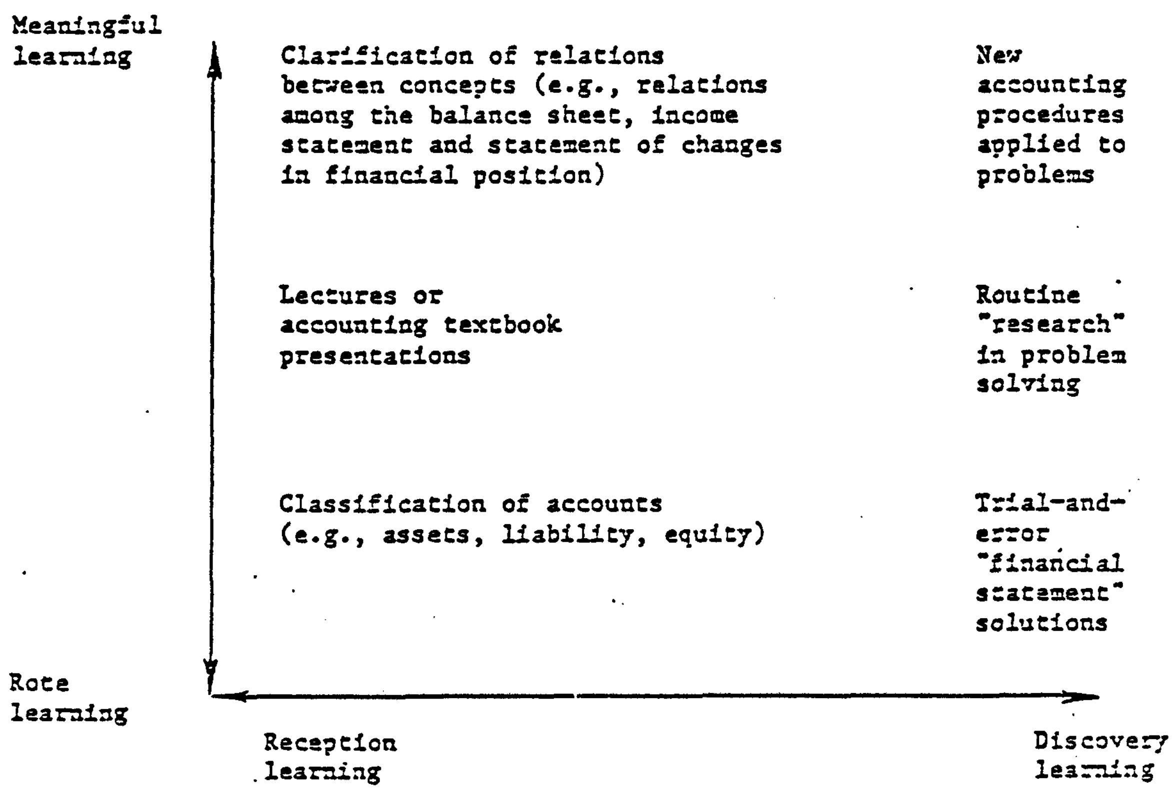
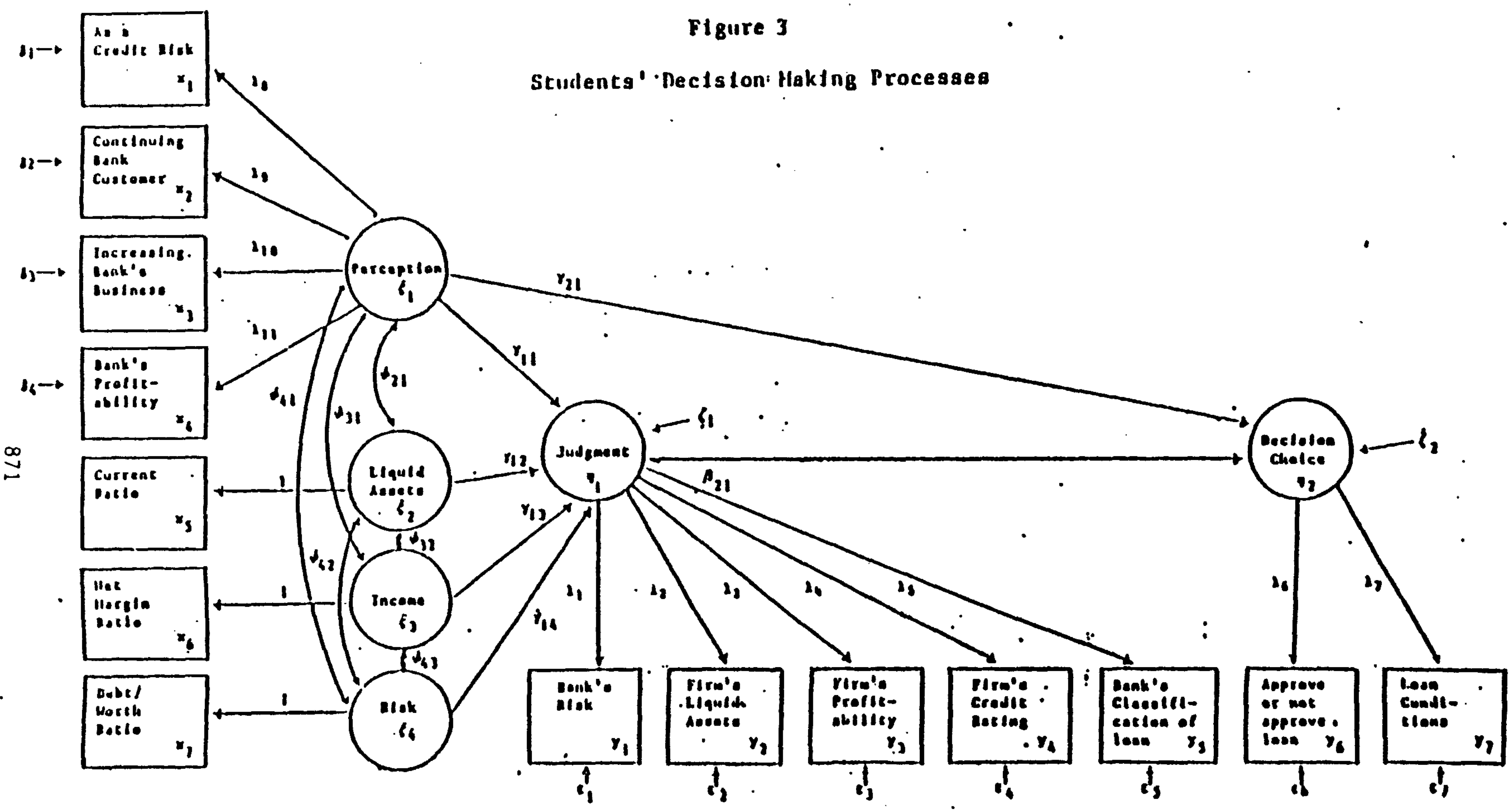


Figure 3

Students' Decision Making Processes



Replacing Paper Users Manuals With On-Line Help

Donna B. Moskow
PRC Inc.

INTRODUCTION

Since 1989, PRC Inc., under terms of a contract with NAVSEA PMS 331,has been responsible for the development and implementation of Maintenance Resource Management System (MRMS) for Intermediate Maintenance Activities (IMAs). MRMS is a computer-based automated information system sponsored by the Assistant Deputy Chief of Naval Operations for Fleet Maintenance and Modernization (OP-43). MRMS is used by IMAs ashore and afloat as a primary management tool providing automated information processing in support of ship and submarine intermediate management. MRMS provides automated information processing in support of a TYCOM Rep component, an IMA component, and a Supply/Financial module.

An integral part of this program is the development and periodic update of users manuals. PRC Inc. is responsible for maintaining all manuals and documentation for MRMS in a paper format with bi-annual updates (called releases) sent to sites around the world. Currently there are 16 users manuals maintained in support of MRMS.

MRMS is an expanding program with additional sites being added on a regular basis. Each site that is brought on board demands individual and unique requirements that must be addressed by MRMS. The addition of sites with their specific requirements requires additional documentation.

The purpose of this paper is to discuss some of the problems associated with the use of users manuals in a paper format. In this respect, PRC is in the process of developing a more effective and efficient tool, namely an on-line user manual. The advantages of such a system are discussed.

PAPER MANUALS DISADVANTAGES

The expansion of MRMS over the last three years has resulted in a corresponding increase in the number of required users manuals as well as a significant increase in the size of each manual. It is not inconceivable that the program will continue to expand and consequently, so will the number and size of the manuals.

The development of a complete users manual, from programming through documentation and on to word processing is a labor intensive process. "It takes on the average three hours to produce a page of software documentation, according to the CO-COMO database of several hundred software development projects." (Singleton,1987) Compounding this are additional software changes after initial documentation is complete, incorporation of changes to contract required Table of Contents, List of Figures, appendices, etc., reproduction requirements, and preparation for shipping.

All of these time measures have an associated cost. "The cost of one labor-year in the mid- 1980's is approximately $100,000, so one hour is worth approximately $50." (Singleton, 1987) When you adjust this figure for inflation and other factors since the 1980's and multiply this by the number of people involved in the production of a users manual, it proves to be quite an expensive process. Another cost associated with paper manuals is reproduction costs. Costs included here are paper costs, binding costs, and copier costs. The larger the manual, the higher the cost is to reproduce it. The last cost associated with paper manuals is the cost of disseminating the manuals to the user. As stated previously, PRC delivers its manuals around the world. An individual site receives an average of 35 manuals. Multiply this by the two release dates per year, and one can see the cost of mailing such a package. On occasion, manuals get lost or damaged in transit, necessitating the need for reproduction and dissemination of replacement manuals, thereby increasing the related costs.

All time considerations and cost associated with the production of MRMS paper users manuals are related to the size of the manuals themselves. Currently the average manual contains 146 double-sided pages. As more sites are added, the size of each manual increases proportionally, to meet the unique needs of the different

sites. Each manual is also bound, by contract to contain a Title Page, Table of Contents, List of Figures, and various appendices, all adding to the size of the manual. Additionally, each manual contains a significant amount of redundant material in that several programs use similar screens and have similar procedures that must be noted in each place they occur. One reason for this is the philosophy that it would make it more difficult for the user to have to flip back and forth to other programs to find procedures. The result of all this is that the manuals have become quite cumbersome to use, bringing into question the utility of the manual for the user.

In order to reduce the time and costs involved with preparation and dissemination of users manuals and to develop more useful instruments, PRC is in the process of developing software that will provide the user with on-line manuals. In this respect, users will have direct and immediate access to the users manuals through the use of a designated function key. In the future, dissemination will be accomplished through the use of disks rather than voluminous paper user manuals. It is necessary that on-line help provide the user with useful, explicit, and correct information. In a sense, it is essential that the on-line help follow the same requirements and standards adapted for the paper manuals.

It should be noted that at this time, PRC's work and research in this area is at a formative stage. Further research is scheduled in the near future to determine specifications such as the hardware platform, the software design, etc.

ON-LINE DOCUMENTATION ADVANTAGES

An obvious advantage to an on-line help system is the reduction and/or elimination of the costs associated with paper users manuals. Significantly effected will be the costs related to reproduction and dissemination of the users manuals. Reproduction requirements will be significantly reduced if not eliminated. It may still be necessary to provide at least one paper copy of the paper manual even under an on-line system. However, even if this is necessary, the reproduction costs would be minimal compared to the current process. Although most of the labor costs associated with development will remain, the labor cost associated with photocopying and mailing will decrease.

The most essential advantages to an on-line system, however, are to the user. PRC is aware that the current users manuals are not being used to their full advantage due, in part, to their considerable size and the amount of repetitive material. An on-line help system will certainly reduce the size of the manuals by eliminating a majority of the redundant information while providing easier access to other screens and information. This system will also provide PRC with an excellent training tool. Currently, PRC conducts training for systems users; however, when that person is transferred to another activity, the replacement may not receive additional training from PRC for some time, if at all.

It has been noted at several sites that first time users of MRMS have difficulty following the users manuals. An on-line system would facilitate the self-training of such individuals. An on-line system will provide a wide variety of examples that show the user how to perform basic database operations like querying and reporting. (Gliedman, 1992) Such a system would be just as beneficial to the beginning user as to the experienced user (assuming comparable computer experience and knowledge of system located at site). Thus, another advantage of an on-line system is its ease of use.

From an information standpoint, the user is assured that the procedures or information he/she receives are of the most current release date. It has been reported that sites do not physically update the manuals the users are currently using. This proves to be a problem when a program's options or functions change and the user is unaware of what is necessary to complete the task at hand and unable to continue without further instruction. The user manual disks will be released in conjunction with the program tapes and will be installed concurrently. Lastly, each site receives a given number of manuals. Some of these manuals are lost, taken by transferring personnel, and even thrown away. This leaves some users without a manual to use. An on-line system would be available to all users at any given time. All these advantages will provide the site user a more effective, current, and easier to use manual.

CONCLUSION

As MRMS has grown over the last three years, the documentation responsibilities have also grown, placing a considerable strain on manpower and budgetary concerns. PRC is also concerned that the users manuals, as they currently exist, are not as effective as is desired. PRC has addressed several disadvantages of the paper system currently adapted; namely, time, cost, and the size of the manuals themselves. These disadvantages affect manpower and budgetary concerns, but more importantly they affect the customer on site. Bruce and Pederson stated that "documentation content and format must satisfy contract requirements and/or applicable company software development policies; it should be appropriate to customer, users, and project needs. Documents should be organized and bound into volumes that are consistent with contract requirements and customer, user, and project needs." Although PRC is currently meeting contract and project requirements, we believe that we are not meeting user needs and requirements. To meet this need and requirement, PRC is in the process of developing an on-line system. The net effect will be a substantial reduction in the costs related to reproduction and dissemination. The most valuable advantage, however, is from the user's standpoint. An on-line documentation system would provide the on site user with an effective, current, attainable, and easy to use manual.

REFERENCES

Bruce, Phillip and Pederson, Sam M. "The Software Development Project - Planning and Management". New York: John Wiley and Sons,1987.

Gliedman, John. "Touching all the bases." PC Sources March 1992 v3 n3 p 449(8)

Singleton, Margaret E. "Automated Code and Documentation Management". New Jersey: Prentice-Hall, Inc., 1987.

Navy-wide Personnel Survey (NPS) 1991: Everyone Has an Opinion[1]

Mary A. Quenette
Organizational Systems Department
Navy Personnel Research and Development Center[2]

The morale and job performance of Navy members take on added importance in an era of down sizing, where each individual must contribute to the increased efficiency required of a reduced force in an unstable world. Navy members' attitudes and opinions represent input vital to the development and continuous improvement of Navy policies and programs; therefore, such opinions must be measured in a systematic and timely fashion, thus furnishing an accurate reflection of the views of its diverse and widespread membership.

The Navy-wide Personnel Survey (NPS), originated in 1990, is an omnibus survey designed to systematically collect opinion data and provide timely information on issues of importance to policy makers. The survey consists of (1) questions that are included on a one-time basis to measure opinions on topics of compelling interest, and (2) questions that are repeated annually to allow the identification and analysis of trends in opinions. Of 230 questions on NPS 1991, 138 questions also appeared on the 1990 NPS.

NPS 1991 questionnaires were mailed in December 1991 to a random sample of 23,821 enlisted personnel and officers with a projected rotation date (PRD) of March 1992 or later. The sampling represented approximately 3 percent of the enlisted population and 11 percent of the officer population. A total of 13,232 surveys were completed and returned for an adjusted return rate of 57 percent.

The survey requested demographic information and measured military members' attitudes and opinions in various areas, including rotation/permanent change of station (PCS) moves, recruiting duty, pay and benefits, training and education programs, quality of life programs, organizational climate, and AIDS education. This paper will focus on a selected topics under the rubric of job issues, including a brief discussion of the differences between the results obtained on the 1990 and 1991 surveys. (Results of the 1991 survey are described in detail in Quenette, 1992; Quenette, et al., 1992a, 1992b; Wilcove & Quenette, 1992a, 1992b. Results of the 1990 NPS are available in Quenette, Kalus, Hase, & Brinderson, 1991a, 1991b, 1991c, & 1991d; Quenette, et al. [in review]).

For statistical analyses, respondents were grouped by paygrade: (1) E-2 and E-3, (2) E-4 through E-6, (3) E-7 through E-9, (4) W-2 through W-4, (5) O-1E through O-3E and O-1 through O-3, and (6) O-4 through O-6. The results presented in this paper are based on data that was weighted by paygrade to reflect each paygrade's actual proportion in the Navy, thereby allowing generalization of sample results to the entire Navy. The margin of error ranged from $\pm$ 01 to $\pm$ 03 percent for the paygrade groups at a confidence level of 95 percent.

Sample Demographics

The demographic characteristics (unweighted) of the respondents were as follows: The majority were male (89% for both enlisted and officers). Sixty-one percent of enlisted were married, with 10

---

[1]Paper presented to the 34th Annual Conference of the Military Testing Association at San Diego, CA, October 1992.

[2]The opinions expressed in this paper are those of the author(s), are not official, and do not necessarily reflect the views of the Navy Department.

percent of married enlisted reporting a military spouse. For officers, 76 percent were married, and 7 percent of married officers had a military spouse. Among enlisted, 66 percent had at least one dependent; for officers the corresponding percent was 78. The mean age was 29.4 years for enlisted and 35.9 years for officers. Among enlisted, racial makeup consisted of 72 percent white, 17 percent black and 12 percent Asian, American Indian, or "Other". Most officers were white (91%); blacks (4%) comprised the second largest group. Seventy-five percent of enlisted did not identify with a specific ethnic group, 13 percent were "Other" ethnic group, and Hispanic and Filipino were 6 percent each. Among officers, 83 percent claimed no specific ethnic identity, 12 percent listed "Other," and 3 percent were Hispanic. Ninety-four percent of enlisted had a high school degree or more and most officers (82%) reported having a bachelor's degree or more.

Job Issues

Career Plans

Among enlisted, 47 percent reported they will stay in the Navy until eligible to retire and 26 percent reported they will leave. A majority of officers (55%) indicated they will stay until eligible to retire, while 14 percent said they will leave. Very few who are eligible to retire indicated that they intend to do so (1% enlisted and 2% officers). In general, the more senior members, both enlisted and officers, were more likely to have made a definite decision to stay. Enlisted subgroups more likely to stay included males, Asians, married personnel, and personnel with dependents. Among officers, females, blacks, married personnel, and personnel with dependents were more likely to stay. Enlisted members who served in the Persian Gulf War were slightly less likely to stay; there was no difference among officers, however.

Ten factors were identified as having an influence on whether or not a member would choose to make the Navy a career (Table 1). The factor receiving the largest negative response was the availability of Navy-sponsored child care. Other important factors included family separations due to duty assignments, amount of sea duty, and family support services. The importance of spouse and family was evident in the response to the question asking whether the member would consider leaving the Navy because of family separations: Approximately one-half of members with dependents answered in the affirmative. Two-thirds of the members would not leave the Navy, however, because of their spouse's career. Retirement pay, retention incentives, and assignment to a high cost area were all listed as negative by 12 percent or less. Half of the questions were answered by subgroups only; for example, the child care questions were answered by members with children. The percentages, therefore, should be interpreted as indicative of importance to specific subgroups, not as importance to the entire sample.

Table 1

Factors Having a Negative Influence on Retention

	Percent Negative'	
	Enlisted	Officers
Navy-sponsored child care*	53	53
Family separations*	51	47
Sea duty*	51	44
Family Support Services	40	41
Living conditions	38	20
Pay	38	26
Spouse career*	16	19
Retirement pay	12	3
Retention incentives	11	4
Assignment to high cost area*	5	2

'Response indicated negative influence on retention.
*Answered by subgroups only.

Duty Assignments

Fifty-seven percent of enlisted personnel claimed to understand the detailing process, but about one-fourth did not. Nearly equal percentages agreed (36%) and disagreed (38%) that the process is fair. Larger percentages of officers understood the detailing process (73%) and agreed that the process is fair (49%) as compared to enlisted. Respondents were also asked to evaluate their current or former detailer on a series of job behaviors. The behaviors were grouped into three general categories, knowledge of job, sensitivity to needs, and customer relations. In general, detailers received the highest ratings on elements of job knowledge. Both enlisted and officers were likely to describe their experience in obtaining their current assignment as "Ran Smoothly" or "Somewhat Smoothly" (69% enlisted and 75% officers), and large majorities reported that they had obtained exactly or nearly the assignment they wanted.

A majority of enlisted personnel (53%) were assigned to sea duty while officers were more likely to be assigned ashore (63%). In general, as paygrade increased, the percentages assigned ashore also increased. Females were far more likely to be serving ashore. Seventy percent of enlisted females had a shore billet and 28 percent were assigned to sea duty, as compared to males with 42 percent ashore and 56 percent at sea. Among officers, 88 percent of females were in shore billets and 10 percent were at sea; 59 percent of male officers were ashore and 36 percent were at sea. Enlisted and officers agreed that 3.0 years is a reasonable length for shore tours, but officers preferred 2.0 years at sea, while enlisted said 3.0 years would be reasonable (medians). Respondents were asked how long they would be willing to extend at sea in order to obtain a shore billet at their home port. Forty-two percent of enlisted and 31 percent of officers were not willing to extend under those circumstances. Of those who were willing, enlisted would extend up to 5 months and officers would extend up to 4.5 months (medians).

Organizational Climate

A series of questions asked respondents to rate their organization on organizational climate issues (Table 2). For enlisted personnel, satisfaction levels hovered around the 50 percent mark for many of the questions. Satisfaction with leadership at the command received lower marks; liking the work and working conditions, and exercising job responsibilities received a higher level of endorsement. Senior enlisted expressed substantially more satisfaction than did the other enlisted paygrade groups. Enlisted females were slightly less satisfied on all questions except when asked if they were glad they chose the Navy, where the percentages were equal. There was no pattern of results for racial groups.

Table 2

Satisfaction With Organizational Climate

	Percent Agreement[a]	
	Enlisted	Officers
Like work I do	70	86
Satisfied with working conditions	63	71
Allowed to exercise job responsibilities	61	77
Satisfied with job	59	75
Enjoy my career	56	80
Chain of Command listens to problems	52	72
Glad I chose Navy	52	74
Satisfied with career development	49	71
Decisions made at appropriate level	45	64
Command support for decisions I make	44	74
Satisfied with quality of leadership	37	65

[a]Percent selecting "Agree" or "Strongly Agree."

Among officers, approximately two-thirds or more expressed satisfaction on all organizational climate questions. Female officers had slightly lower levels of satisfaction: For most questions, the male/female difference was five to six percentage points or less. There were slight differences by race for a few questions, with whites and Asians a few percentage points higher in satisfaction. Senior officers and Warrant Officers were slightly more satisfied than junior officers on a few questions.

Equal Opportunity

Enlisted personnel responded positively to most of the equal opportunity questions (Table 3). About three-quarters agreed that they are treated fairly by their supervisor, their CO and XO support equal opportunity, and their work assignments are fair. Agreement was well over 50 percent for the remainder of the questions. By paygrade, agreement increased with paygrade level for all questions. As with the previous set of organizational climate questions, females were slightly less in agreement than males.

Table 3

Satisfaction With Equal Opportunity

	Percent Agreement[a]	
	Enlisted	Officers
Immediate supervisor treats me fairly	77	90
CO actively supports EO	76	88
XO actively supports EO	73	87
Work assignments are fair	72	89
Efforts to improve EO in Navy	62	81
Chain of Command effective in resolving EO problems	60	78
At Captain's Mast, I would be treated fairly	59	79

[a]Percent selecting "Agree" or "Strongly Agree."

For officers, agreement was nearly unanimous for the fairness of treatment from supervisors, CO and XO support for equal opportunity, and the fairness of work assignments. Agreement for the remaining questions was around 80 percent. By paygrade, junior officers agreed less on a few of the questions. Once again, females agreed less than males. Results by race were mixed, but blacks expressed less agreement on a few questions.

Enlisted were more favorable to women serving aboard combat ships and combat aircraft than to women aboard submarines. There were trivial differences by paygrade, sex, educational level, marital status, or deployment to the Persian Gulf. Overall, officers were less favorable than enlisted on such assignments for women. Paygrade and marital status differences were slight, and results for educational level were mixed. Female officers agreed at a much higher level, about 35 percentage points higher than males, on all three questions, and there was a slight tendency for officers who had been deployed for Operation Desert Shield/Storm to agree less as compared to those who were not deployed.

Sexual Harassment

Approximately three-quarters of enlisted and officer respondents had received training in the prevention of sexual harassment in the past 12 months. Respondents were asked a series of questions about the types and frequency of sexual harassment behaviors directed at them during the preceding 12 months, the persons who engaged in such behaviors, and whether or not they had been the victim of sexual assault or rape. As seen in Table 4, there were very large differences by gender; 10 percent or less

of enlisted males said they had been the target of such behaviors, while the percentage of female enlisted who responded likewise ranged from 12 percent to 58 percent. By paygrade for enlisted, the percent who indicated that they had been subjected to the behaviors decreased as paygrade increased. The most common behavior was teasing, jokes, remarks, or questions, and the least common behavior was pressure for sexual favors.

Table 4

Percent* Reporting Sexual Harassment Behaviors

	Enlisted		Officers	
	Females	Males	Females	Males
Teasing, jokes, remarks	58	10	43	4
Looks, staring, gestures	55	8	33	2
Whistles, calls, hoots	50	5	28	1
Touching, leaning over, cornering	35	6	16	2
Pressure for dates	31	3	10	1
Letters, phone calls, sexual materials	18	3	10	2
Pressure for sexual favors	12	2	3	0

*Percent selecting "Once" or more in frequency.

Among officers, the percent who indicated that they had been the target of the behaviors listed was somewhat smaller than for enlisted. Again, large differences emerged when the data were broken down by gender: Four percent or less of male officers said they had been subjected to the behaviors listed, while female officers' responses ranged between 3 percent and 43 percent. As with enlisted, the most common sexual harassment behavior was teasing, jokes, remarks or questions, and the least common was pressure for sexual favors. The overall rate of sexual harassment, defined as the percent who had experienced one or more of the behaviors on at least one occasion during the preceding 12 months, was calculated for each gender group for enlisted and officers. The overall rates were 73 percent of enlisted females, 18 percent of enlisted males, 57 percent of female officers, and 8 percent of male officers.

Enlisted reported that harassment came most frequently from coworkers and "Other," and from military enlisted. Enlisted males reported that females were the most frequent harassers (54%), followed by male harassers (31%). A few (14%) reported that they had been harassed by both sexes. Enlisted females were harassed almost exclusively by males (95%). Officers indicated that most harassment came from coworkers, followed by "Other," and subordinates, and from military officers, followed by military enlisted. Among officers, males were harassed primarily by females (74%) and females were harassed almost entirely by males (97%). Finally, 5 percent of enlisted females, 2 percent of enlisted males, and 1 percent of officers, both males and females, reported that they had been the victim of rape or sexual assault during the preceding 12 months.

Discussion

Since the 1990 survey, the percent who indicated that they *definitely* intend to stay in the Navy until eligible for retirement increased 5 percent for officers and 3 percent for enlisted. There was a very slight decline in the percentages who said they *probably* would stay. There was also a slight decrease in percentages who said they plan to leave.

The rates of agreement with the organizational climate, equal opportunity, and sexual harassment questions remained relatively stable since the 1990 survey. For the organizational climate questions, the only change of note was that enlisted were slightly more satisfied with the quality of leadership at their command this year. Neither officers nor enlisted appear to have had a significant change of opinion since last year regarding equal opportunity. There was a slight decline since last year in the rate of three of the

sexual harassment behaviors aimed at female officers, but no meaningful change for male officers or for enlisted personnel of either gender.

Differences between the 1990 and 1991 results must be interpreted cautiously, since most of the differences were small. Some differences could be a result of sampling error or other unidentified sources of variability. Most of the questions discussed in this paper will appear in the 1992 NPS, and the third data point will allow the identification and examination of trends. New topics appearing in the 1992 NPS, scheduled to be mailed in November, include Navy uniforms, Navy Exchanges, skill training, shipboard recreation, Navy Core Values, health promotion programs, and drug and alcohol programs.

References

Quenette, M. A. (1992). *Navy-wide Personnel Survey (NPS) 1991: Report of findings* (NPRDC TR-92-20). San Diego: Navy Personnel Research and Development Center.

Quenette, M. A., Baker, G. H., Culbertson, A., Magnusson, P., Rosenfeld, P., Thomas, M., & Tyburski, D. (in review). *Navy Personnel Survey (NPS) 1990: Report of findings.* San Diego: Navy Personnel Research and Development Center.

Quenette, M. A., Gordon-Espe, M., Eliassen, D., Kalus, S., Hase, J., & Brinderson, C. (1992a). *Navy-wide Personnel Survey (NPS) 1991, graphic presentation of results for enlisted personnel* (NPRDC TN-92-20). San Diego: Navy Personnel Research and Development Center.

Quenette, M. A., Gordon-Espe, M., Eliassen, D., Kalus, S., Hase, J., & Brinderson, C. (1992b). *Navy-wide Personnel Survey (NPS) 1991, graphic presentation of results for officers* (NPRDC TN 92-21). San Diego: Navy Personnel Research and Development Center.

Quenette, M. A., Kalus, S., Hase, J., & Brinderson, C. (1991a). *Navy Personnel Survey (NPS) 1990 survey report, statistical tables (enlisted personnel)* (NPRDC TN-91-17 Volume 1). San Diego: Navy Personnel Research and Development Center.

Quenette, M. A., Kalus, S., Hase, J., & Brinderson, C. (1991b). *Navy Personnel Survey (NPS) 1990 survey report, statistical tables (officer personnel)* (NPRDC TN-91-17 Volume 2). San Diego: Navy Personnel Research and Development Center.

Quenette, M. A., Kalus, S., Hase, J., & Brinderson, C. (1991c). *Navy Personnel Survey (NPS) 1990 survey report, graphical representations (enlisted personnel)* (NPRDC TN-91-17 Volume 3). San Diego: Navy Personnel Research and Development Center.

Quenette, M. A., Kalus, S., Hase, J., & Brinderson, C. (1991d). *Navy Personnel Survey (NPS) 1990 survey report, graphical representations (officer personnel)* (NPRDC TN-91-17 Volume 4). San Diego: Navy Personnel Research and Development Center.

Wilcove, G. L. & Quenette, M. A. (1992a). *Navy-wide Personnel Survey (NPS) 1991, statistical tables for enlisted personnel* (NPRDC TN-92-22). San Diego: Navy Personnel Research and Development Center.

Wilcove, G. L. & Quenette, M. A. (1992b). *Navy-wide Personnel Survey (NPS) 1991, statistical tables for officer personnel* (NPRDC TN-92-23). San Diego: Navy Personnel Research and Development Center.

EMPATHIZING WITH THE SURVEY RESPONDENT[1,2]

Gerry L. Wilcove
Navy Personnel Research and Development Center

Background

Vice Admiral J. M. Boorda, the Chief of Naval Personnel, commissioned the Navy Personnel Survey (NPS) 1990. The NPS is designed to be an annual comprehensive survey composed of permanent items, that permit the detection of trends, and topical items that may vary from year to year. The 1990 NPS addressed a variety of areas, including rotation/permanent change-of-station (PCS) moves; recruiting duty; pay and benefits; training and education; quality-of-life programs concerned with family support services, child care, and housing and recreational services; organizational climate, including equal opportunity and sexual harassment; and education about Acquired Immune Deficiency Syndrome (AIDS).

NPS 1990 was composed of both multiple-choice items and sections that permitted respondents to submit written comments. With multiple-choice items, the opinions of thousands of personnel can be sampled, and generalizations can be drawn based on statistical analysis. This statistical approach, however, limits our ability to empathize with the individual officer or enlisted person. Written comments, on the other hand, facilitate empathy and understanding because they describe the wealth of experiences and emotions that cannot be captured by statistics. Put another way, written comments ". . . give you a feeling for people you don't know personally" (Plummer, 1974).

Method

The comments from 450 individuals were selected for analysis. More specifically, comments from 50 individuals were randomly selected for each of the nine areas covered by the NPS (such as pay and benefits). Comments were combined into categories for each of the areas, and the number of comments in each category were counted. Categories with the greatest counts received the most attention in the published report (Wilcove, 1991). Comments were included from both officers and enlisted personnel, males and females, and grades of all levels.

---

[1]The opinions expressed expressed in this paper are those of the author, are not official, and do not necessarily reflect the views of the Navy Department. The study on which this paper is based was performed under reimbursable work unit 981WRB1007. Dr. Wilcove can be reached at DSN 553-9120 or (619) 553-9120.

[2]This paper was presented at the 34th Annual Conference of the Military Testing Association on October 29, 1992, San Diego, CA.

Purpose

The purpose of Wilcove (1991) was to allow policy makers and managers to empathize with personnel by learning more about their objective circumstances and their subjective reactions. The purpose of the present paper is to extrapolate from written comments to broad concepts in an attempt to understand the behavior of large groups of personnel. If these and other explanatory concepts are consensually and empirically validated, then the Navy will be in a better position to predict the behavior of selected personnel in the areas of retention, performance, and education. The Navy will also be able to better structure the environments of personnel to maximize their growth and satisfaction.

Let us now turn to the written comments themselves and the broad concepts they stimulated.

Findings

The first comment by an E-2 shows how concepts about equity serve as determinants of a person's perceptions and level of satisfaction:

"As it stands I'm forced to live on board while my civilian friends are making almost double and in some cases triple. Tell me that it makes sense to serve your country and people in civilian life are doubling what you make. Not only that, but I'm the one willing to die for America and I can't even afford to live off ship."

Another way of looking at this comment is that the person believes that extrinsic rewards, in this case--money, should be commensurate with the value of his motives, in this case--patriotism.

However, even when equity is a determinant of an individual's perceptions, it is not always the only concern as suggested by the following comment:

"The Surgeon General of the Navy has done all he could to support us physicians relevant to pay and salary compared to the civilian community. His effort has been applauded by a majority of the Navy Physicians and we respect him for it. For myself, I can live with the current salary structure."

In short, when one feels valued by the organization, equity becomes a more comprehensive concept than simple comparisons of dollars and cents.

The next comment increases our appreciation for personnel with financial concerns whose horizons are not restricted to the present:

"Three years shore time does not allow a young family the time to become financially secure enough for the future. Buying a house, for example, is very important, but to make a profit you must live in the house for at least five years." (E-4)

The next comment illustrates both the complexity of situations and the human experience:

"Recruiting duty has been my most challenging tour. There is a lot of stress involved and a lot of long days. I take pride in my job and I've done my best out here. I think the tour should be cut down to 2 years. I think recruiting should be strictly volunteer, not being forced or 'nominated' for recruiting. It has been a learning experience and I think it will help me be a better leader/manager in the fleet. (E-6)

Notice the multidimensional nature of the recruiting experience for this individual. It is both challenging and stressful. While the individual has adapted to conditions, and taken pride in the job they've done, they recommend that certain changes be made. However, even with the job's present parameters, they feel it will make them a better manager and leader.

A comment by a commander suggests that empathy with an organization's position, even when that position is based on cost-benefit considerations, goes only so far in determining an individual's opinion:

"The GI Montgomery Bill now is very helpful. (I'm) using it towards a masters degree. The reimbursement plan, however, is unsatisfactory. When you register for courses, you must pay for the entire semester up front. For me, that is about $1500 a semester. Repayment is made, then, month by month. I can afford it, but I think about those who can't make the upfront payment and then get partially reimbursed over up to five months. I know this deters some people from using their educational benefits and who never get started. I guess this system is to ensure no one cheats the system, but it is having an undesirable effect on non-cheaters."

The individual understands the basis for the organization's policy, but believes it to be an ill-advised approach to the administration of educational funds.

An E-6's comment suggests that if conducive environmental conditions exist, some individuals will be able to reach out for help in their time of need:

"I've utilized the counseling program once; confidentiality was respected, not like rumor has it. Professional personnel on staff made it more comfortable to communicate with. I had the impression the program was run by volunteers (i.e., Navy wives, retired personnel, etc.)."

This comment suggests a relationship between inaction and negative stereotypes. Perceiving that counselors do not respect confidentiality would lead to inaction.

The diversity of human reactions is underscored by the following two comments about child care in the Navy:

"We have found the child-care workers to be quality people with a real concern for the children." (E-5)

"Professional child-care facilities, including the Navy's, are 'cattle barns' for children. My children will never experience them." (O-3)

Alderfer (1969) from Yale postulated 3 classes of human needs: existence, relatedness, and growth. Existence needs refer to a person's physical and physiological requirements, including the need for "creature comforts," as the following comment graphically illustrates:

"There is a shortage of affordable rentals in the Charleston, SC, area. I am a chief and I live in a very small three bedroom house. My BAQ [Bachelor's Allowance for Quarters] and VHA [Veteran's Housing Allowance] are not sufficient to cover my rent and utilities. I use the air conditioner and heat very sparingly to keep my electric bills down. I set my thermostat on 84 degrees in the summer and 64 degrees in the winter, both of which are not very comfortable." (E-7)

The following comment by an E-5 portrays the dynamic interplay among all 3 classes of human needs postulated by Alderfer:

"I am very happy with my base housing [note the implications here for satisfying his existence needs]. Base housing has had a very positive effect on my job performance [one can see here how fulfilling one's existence needs can free an individual to grow on the job]. I am very thankful for the Navy improving the quality of life for me and my family [this statement reflects the social connection between the individual and the organization]."

The final two comments remind us of the proactivity of individuals. Individuals have the capacity to impact their environment and are not simply reactive to it. This capacity is reflected in the first recommendation for change offered by a lieutenant commander.

"The development of mass transit to and from housing areas would reduce traffic congestion, pollution, parking requirements. Some of the costs could be borne by riders and schedules could be developed to support military work hours. Possible reduction of U/A's, accidents, etc., would increase productivity."

The second recommendation by an E-7 raises the issue of conflict resolution and, more particularly, the role of communication in the reduction of interpersonal and intergroup conflict:

I believe there should be a forum for discussing racial, ethnic, and gender differences. There is a lot of ambiguity on what racial and ethnic slurs, omissions, and the such are, along with the fine lines between sexual harassment, politics, and discrimination. If there is no forum for personnel to vocalize their thoughts, feelings, and fears, then there tends to be a lot of resentments held in. Let's bring everything out in the open, discuss it and resolve any differences there might be."

DISCUSSION

Reports that focus on written comments are typically popular and meaningful to policy makers and managers, and Wilcove (1991) was no exception. The purpose of that report was to allow individuals to empathize with the survey respondents. However, written comments, much like letters to the editor, tend to be critical in tone. This tendency raises the question of representativeness--the extent to which comments

represent all personnel and the extent to which a comment given by an individual represents their overall attitude.

I am currently conducting research on the issue of representativeness and other issues related to the usefulness of written comments. Five "usefulness" criteria are being examined. First, to what extent do comments describe favorable situations in the Navy? Managers and policy makers need to have such feedback as part of their evaluation regarding which policies and procedures to maintain. In short, comments are most useful when they give a balanced picture of Navy experiences. Second, to what extent do comments elaborate on responses to yes-no and multiple choice survey items? Elaboration provides depth and understanding in specific areas, and, across items, helps to construct an overall picture of Navy experience. Third, to what extent do comments identify positive and negative situations untapped by survey items? Fourth, to what extent do personnel offer recommendations in their comments for improving the way the Navy prepares for defense, treats its personnel, and spends its money? Fifth, to what extent do comments promote empathy for the individual enlisted person and officer by conveying their feelings, values, and opinions and by elaborating the situations and events that determine the very fabric of their lives and careers? Having determined the answers to these questions in ex post facto research, I will then determine experimentally how well different survey formats can optimize the usefulness of written comments in the five criterion areas.

REFERENCES

Alderfer, C. P. (1969). Existence, relatedness, and growth: Human needs in organizational settings. New York: The Free Press, 1972.

Plummer, J. T. (1974). Application of life style research to the creation of advertising campaigns. In W. D. Wells (Ed.). Lifestyle and psychographics. Chicago: American Marketing Association.

Wilcove, G. W. (1991). The Chief of Naval Personnel asked, and here is what they said! (An analysis of written comments from the Navy Personnel Survey 1990) San Diego: Navy Personnel Research and Development Center.

LET'S HEAR FROM THE RESERVES: RESULTS OF THE 1991 NAVAL RESERVE SURVEY

Herbert George Baker, PhD
Navy Personnel Research and Development Center
San Diego, CA 92152-6800

ABSTRACT

The 1991 Naval Reserve Survey was launched by the Chief of Naval Personnel to assess the attitudes and opinions of Naval Reservists with respect to overall reserve experiences, and particularly with respect to Operations Desert Shield and Desert Storm (DS/S). The survey was mailed to a sample of reservists (N=31,763) in November and December of 1991. The adjusted response rate was 44 percent (N=12,231). Questionnaires elicited Social security Numbers (SSN), in order to enable drawing demographic data from the reserve master tape. Of the total respondents, there was an 89 percent match with SSN (N=10,881).

The survey was in three parts: (1) a section for all respondents, dealing primarily with overall satisfaction with reserve administration, opportunities, etc.; (2) a section for reservists recalled to active duty, addressing in- and outprocessing, administrative concerns, reception by active duty units, and impact on family, job, and finances; (3) a section for personnel who had been released from active duty after their call-up. Each section also included space for write-in responses.

Respondents were generally satisfied with their overall reserve service, and with their experiences during DS/S. Specific problem areas were identified, and their implications for policy and practice are discussed. Also included are selected sub-group comparisons (e.g., doctors, nurses, medical enlisted, etc.) on key items (loss of income due to call-up, threat to job, business, or practice, etc.).

Introduction

Periodic assessment of member attitudes, opinions, and satisfaction is a requirement of organizational life. Such information provides a snapshot of the current organization, in terms of how well its central processes are functioning and in terms of the morale of its people. Systematic and scientifically defensible data collection is a requisite for providing timely and accurate information to the leaders of the Naval Reserve. This is particularly true following reserve personnel call-ups, such as occurred in support of the recent Operations Desert Shield and Desert Storm (DS/S).

The 1991 Naval Reserve Survey was launched to collect information on attitudes, experiences, and satisfaction of Naval Reservists, particularly with respect to DS/S, to assist the Navy leadership in enhancing ongoing reserve administration as well as mobilization policies and practices.

Approach

Questionnaires were mailed to 31,763 Naval Reservists during November and December 1991. Addresses were provided by Naval Reserve Headquarters for 10 percent of all reservists who were not recalled for DS/S, 25 percent of recalled reservists serving in medical occupations, and 100 percent of recalled reservists serving in non-medical occupations.

The Naval Reserve Survey requested the respondent's Social Security Number (Item 1). Subsequent to their serialization and optical scanning, questionnaires were matched with personal data on the reserve master tape. This allowed acquisition of personal data without elicitation in the questionnaire itself. In addition, respondents were asked to indicate the dates of significant events in their mobilization. From these responses were calculated certain information such as time between notification and entry on active duty. Respondents were also invited to provide write-in responses in each section of the survey. These written responses (from an estimated 4000+ respondents) have not yet been content analyzed.

Reminder/Thank You postcards were mailed to all questionnaire recipients approximately four weeks after the questionnaire mailout. Only those surveys received on or before 31 March 1992 were included in the data analyses. After subtracting for surveys which were undeliverable due to faulty addresses, the adjusted response rate for the survey was 44 percent. There was an 89 percent successful match between returned questionnaires and the reserve master tape (i.e., SSN matchup).

Results

Sample Characteristics

More than two-thirds (70%) of the respondents were enlisted personnel, and slightly more than one-fourth (28%) were in medical occupations. Women comprised 21 percent of the sample.

The sample was fairly evenly split among those not recalled to active duty (35%), those recalled and assigned to a base in continental U.S. (CONUS) (38%), and those who were assigned to a forward area (27%).

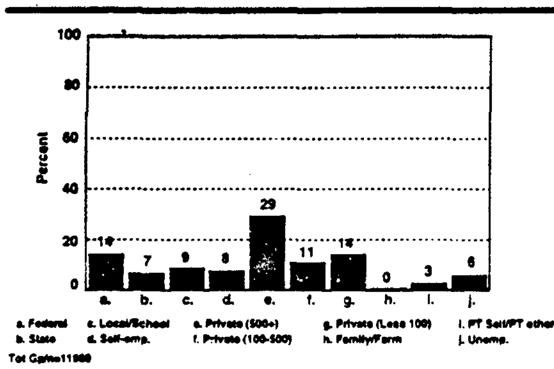
Seventy-seven percent of the respondents were over 30 years, 40 percent over 40. Among other implications, the data indicate that reservists have developed relatively stable career patterns, and more complex family situations, both of which interact with mobilization concerns.

The highest racial presence was that of Black, which, at three percent for officers and eight percent for enlisted, was well below their representation in society at large. Similarly, claimed ethnic background was negligible, with more than 80 percent of both officers and enlisted being category Y (no ethnic background claimed).

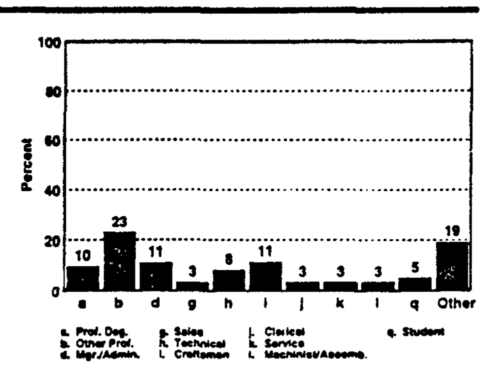
Figures 1 and 2 reflect information about the civilian jobs of the respondents. Twenty-one percent were employed by state or federal governmental agencies. By far the largest single group was those employed by larger private firms having 500 or more employees. Only six percent were unemployed.

The high skill profile of Naval Reservists is again evident, with over half of the respondents (52%) being in the occupational categories of professional degree, other professional, managerial/administrative, or technical. Another 11 percent were craftsmen. The highly skilled nature of the reserve population speaks again to the need for meaningful work, as discussed earlier.

Civilian Employer (Q4)

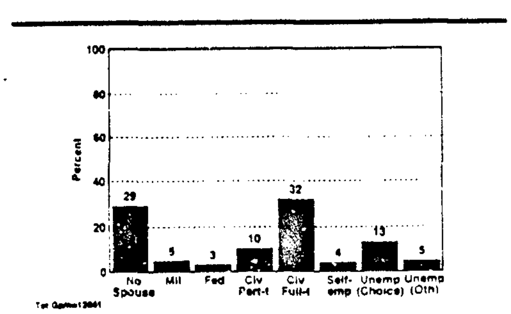


Civilian Job (Q5)



Only seven percent of the respondents were not working at the time of their call-up. With respect to the work situation of the respondent's spouse (Figure 3), of those with spouses, only 18 percent reported their spouses to be unemployed either by choice or involuntarily. If the 29 percent of the respondents having no spouse are not considered, the proportion of working spouses is even higher. The obvious indication is that the families of most Naval Reservists, like those of the majority of Americans, depend on two wage earners. This situation should be considered in reserve policy, reserve management, and reserve mobilization. In only a few cases (3%) were reservists' spouses also in the reserves. Although the numbers are small, for those few families where both spouses are activated, the impact on parenting, family stability, and finances would be considerable.

Spouse's Job (Q7)

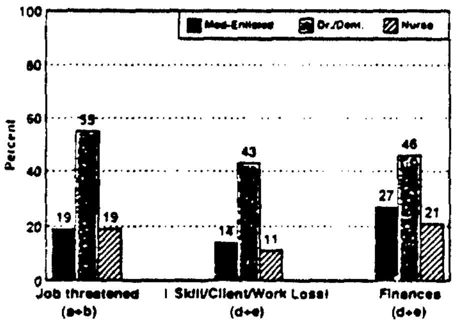


The Costs of the Gulf War

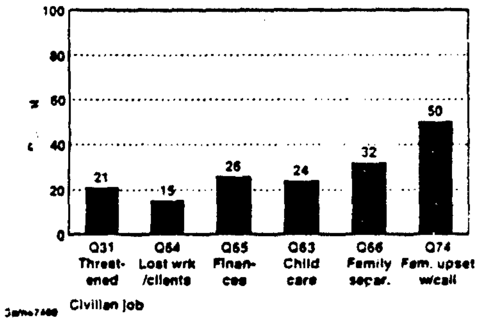
A glance at Figure 4 shows the dramatic differences between the doctor and dentist occupations and others. Over half of the medical occupations felt their jobs/practices were threatened by their being called to active duty. And more than 40 percent of them felt they sufferred finacial loss as well as a loss of skills, clients, or work. It is important to note, however, that nearly one-fifth of the medical enlisted and the nurses perceived a threat to their jobs also; and, large numbers of these individuals also indicated a financial loss due to their activation.

Some sense of the financial impact of Desert Shield/Storm call-up on reserve members can be gained from the information in Figure 5. Many reservists felt their jobs or practices to be threated by their being called to active duty, many lost work or clients. Financial loss was felt by 26 percent of the respondents, and nearly as many had problems with child care.

Impact Of Call-up On Civilian Job Concerns (Q31,64,65)



Civilian Job and Finances - Negatives

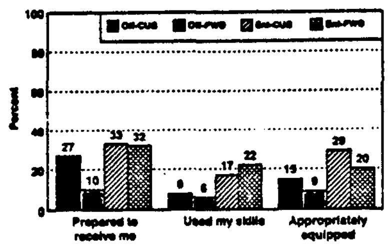


Opinions About the Recall Experience

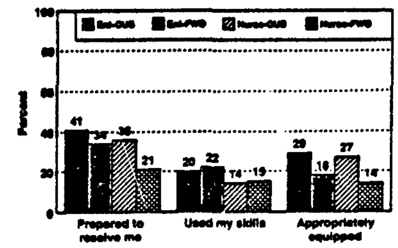
Figures 6 and 7 show that, of both active duty subgroups (male non-medical and medical), there is consistently more dissatisfaction with bases in CONUS than with forward area bases. This applies similarly to perceptions that the base was prepared to receive them, that their skills were properly utilized, and that the bases were appropriately equipped.

Figure 8 shows the levels of dissatisfaction with three elements of the recall. Fifteen percent were dissatisfied with inprocessing. Nearly one-fifth rated PSA staff as not helpful or not knowledgeable. And their duty stations were often not properly equipped (23%), or not prepared to receive the incoming reserve members (32%).

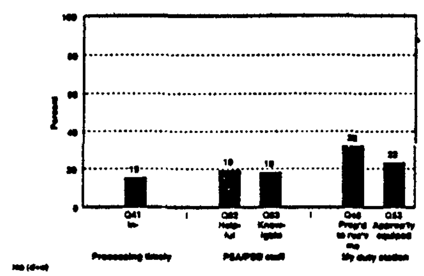
Dissatisfaction With Duty Station (Q46,48,53)
ACDU Men Nonmedical Occupations



Dissatisfaction With Duty Station (Q46,48,53) ACDU Medical



In/Out Processing and Duty Station – Negatives



Despite the presence of a number of dissatisfiers, ninety-five percent were pleased and proud to serve. Somewhat less overwhelming but still impressive is the agreement with being enthusiastic about being called-up (66%). More than three-fourths of the respondents agreed that their overall recall experience and their duty assignments while on active duty were satisfactory. And 76 percent agreed that there was good local community support.

Overall Satisfaction with Reserve Service

A number of items having to do with the reserve program in general are addressed by Figure 9. Fifteen percent of the respondents indicated their intent to leave the reserves. In addition, from one-fifth to one-third of the sample voiced dissatisfaction with military job opportunities, opportunities for training and education, reserve administration, and training for recall.

Reserve Program – Negatives

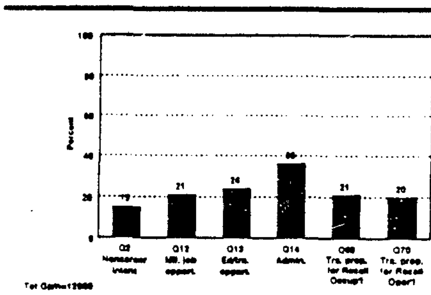
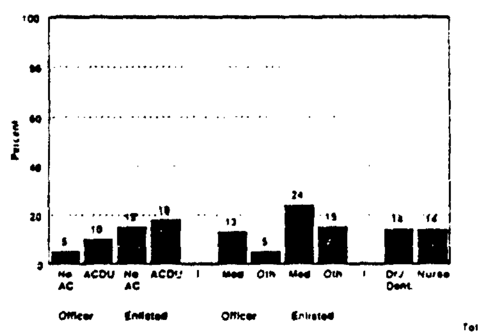


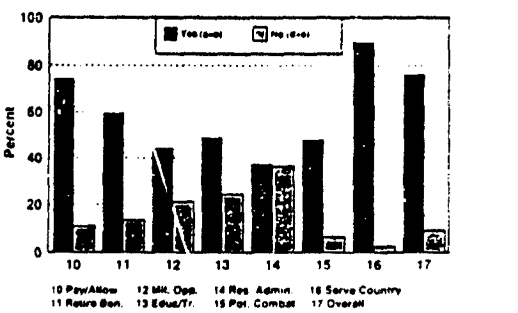
Figure 10 compares the percentages of those respondents intending to leave the reserves across several subgroups. On the left, officers and enlisted, those called-up and those not. A higher percentage of enlisted intend to leave than officers, and, in both cases, those having been brought on active duty are more likely to leave. In the center section of the graph, it can be seen that medical occupation personnel are more likely to leave than non-medical; again, the enlisted of both subgroups are more likely to leave than the officers. Finally, on the right are shown the non-career intentions of doctors and nurses. Despite voicing concerns in several areas, both doctors and nurses, at 14 percent, approximate the total sample in their intentions to remain or leave.

Figure 11 illustrates satisfaction and dissatisfaction with a variety of features. Dissatisfaction edges higher with respect to military opportunities and for opportunities in the reserves for education and training. Dissatisfaction is more pronounced, and nearly equals satisfaction with respect to reserve administration. However, overall satisfaction is very high, and there is overwhelming satisfaction with reserve service for opportunity to serve one's country.

Reserve Non-Career Intent (Q2)
Get Out Or Go IRR (c+d)



Satisfaction with Reserves
(Drill Status) (Q10-17)



Tot Grp n = 12231

Despite some significant areas of concern, the "big picture" is gratifying. More than 70 percent said they were generally satisfied with the reserves prior to DS/S. Furthermore, of those in the sample who were recalled to active duty, 78 percent said that, overall, their recall experience was satisfactory. Finally, there was overwhelming agreement that they were proud to serve their country during Operations Desert Shield and Desert Storm (95%).

CONCLUSIONS

1. Most reservists are satisfied with their overall reserve experiences.

2. There are areas of concern (e.g., reserve administration, mobilization in-processing and skills utilization) that need addressing to improve mobilization effectiveness..

3. Despite personal costs and inconveniences, Naval reservists were ready, pleased, and proud to serve.

The Factor Structure of Cognitive, Spatial, Perceptual, and Psychomotor Tests[1]

Norman G. Peterson and Rodney L. Rosse
American Institutes for Research

Teresa L. Russell
Human Resources Research Organization

Introduction

This paper reports an investigation of the factor structure of the Armed Services Vocational Aptitude Battery (ASVAB) and spatial, perceptual, and psychomotor tests from the experimental test battery developed under the auspices of the U. S. Army's Project A and Career Force projects (Peterson, Hough, et al., 1990; Peterson, Russell, et al., 1990). In particular, we report on the extent to which the test scores appear to reflect a general ability factor, often referred to as "g", and more specific ability factors.

There is a long history regarding the concepts, measurement and utility of application of general cognitive ability and/or multiple abilities (Spearman, 1927; Thurstone, 1938; Brogden, 1951; Jensen, 1980). In this paper we include measures of abilities that are have been subjected to considerable research in the military sector and are under serious consideration for possible implementation in operational selection and classification settings. Some of the measures are not ordinarily considered cognitive measures, but we have included them because they are intended to supplement cognitive tests currently used in military selection and classification.

Ree and colleagues have recently completed research on "g", especially with regard to the ASVAB (Ree & Earles, 1990, 1991a, 1991b; Ree, Earles, & Teachout, 1992). We compare our findings with some of theirs, particularly with regard to the pattern of ASVAB subtest loadings.

Method

Sample. Data were collected for a sample of approximately 50,000 persons entering the U.S. Army during the period from mid-1986 to mid-1987. These persons were enlisting in one of nineteen Military Occupational Specialties (MOS) chosen to be representative of all enlisted Army jobs, to have large numbers of incumbents, and to be high-priority MOS in event of a national emergency. For purposes of these analyses, a subsample of 6,436 persons was randomly selected. The sample was approximately 68% white, 25% black, 3.5% Hispanic, and 3.5% "other". About 12% were female. (Full description of the sample can be found in Campbell & Harris, 1990 and Peterson, Russell, et al., 1990.)

Measures. We used thirty-three scores from the ten ASVAB subtests, six paper-and-pencil tests of spatial ability, and ten computer-administered tests of reaction time, memory, perception, and psychomotor ability. The tests and scores are shown in Table 1. Full details on the derivation of the scores and their psychometric properties can be found in Peterson, Russell, et al. (1990). The ASVAB scores were scores of record, therefore, they were generally administered prior to the soldiers' entry into service. The experimental battery scores were all administered within the first three days of the soldiers' arrival at a reception battalion as they were entering service. Note that the first six computer-administered tests use the reaction time format and all use decision time (the time between appearance of a stimulus and a decision by the examinee) and percent correct scores. The Movement Time score uses the pooled movement time for these tests, except Number Memory. Movement time is the time it takes for an examinee to make a motor response indicating their answer <u>after</u> they have made a decision about the correct answer.

[1]This research was funded by the U.S. Army Research Institute for the Behavioral and Social Sciences, Contract Nos. MDA903-82-C-0531 and MDA903-89-C-0202. All statements expressed in this paper are those of the authors and do not necessarily reflect the official opinions or policies of the U.S. Army Research Institute or the Department of the Army.

Table 1. Thirty-three scores from ASVAB and Project A Experimental Battery submitted to Principal Components Factor Analysis.

Armed Services Vocational Aptitude Battery Subtest Scores
- General Science
- Word Knowledge
- Electronics Information
- Paragraph Comprehension
- Mechanical Comprehension
- Arithmetic Reasoning
- Mathematics Knowledge
- Number Operations
- Coding Speed
- Auto/Shop

Paper-and-Pencil Spatial Tests: Number Correct Scores
- Assembling Objects
- Reasoning Test
- Maze Test
- Object Rotation Test
- Orientation Test
- Map Test

Computer-Administered Tests
- Simple Reaction
  - Decision Time
  - Percent Correct
- Choice Reaction
  - Decision Time
  - Percent Correct
- Perceptual Speed/Accuracy
  - Decision Time
  - Percent Correct
- Target Identification
  - Decision Time
  - Percent Correct
- Number Memory
  - Operations Decision Time
  - Percent Correct
- Short-Term Memory
  - Decision Time
  - Percent Correct
- Movement Time Pooled
  - Mean of Movement Times
- Target Tracking 1 (One-hand Tracking)
  - Mean Log (Distance + 1)
- Target Tracking 2 (Two-hand Tracking)
  - Mean Log (Distance + 1)
- Target Shoot (Dist)
  - Mean Log (Distance + 1)
- Cannon Shoot
  - Mean Time Discrepancy

The last four computer-administered tests measure perceptual ability involving movement and psychomotor ability. The two tracking tests use a distance error score, i.e., the averaged distance between a crosshair and the center of a target the examinee is attempting to track. The Target Shoot test likewise uses a distance error score, but it is the distance between a crosshair and a target at the instant the examinee "shoots" at the moving target. The Cannon Shoot test uses a time error score, indicating the degree to which an examinee is "off" in timing the firing of a cannon "shell" at a moving target. As noted in the *Introduction*, all of these measures involve cognitive ability, but clearly the last four computer-administered measures (the tracking and shooting measures) are intended to measure psychomotor ability and might not traditionally be included in a cognitive ability battery.

Method. The thirty-three scores were correlated and the correlation matrix was submitted to principal components analysis. We performed a parallel analysis to determine the number of factors that could be reliably extracted, given the number of scores and sample size (Allen & Hubbard, 1986). This analysis indicated that thirteen factors should be extracted.

Results and Discussion

Table 2 shows the factor loadings for the first thirteen principal components. The negative loadings should be ignored. They are due to the fact that higher values for the computer-administered scores for distance and decision time connote "worse" performance, causing inverse correlations with the paper-and-pencil test scores. Note that the cumulative eigenvalue for this solution is 24 which is 73% of the total variance for the thirty-three scores. The eigenvalue for the first factor, usually taken as the measure of "g" for the battery, is 7.8 which is about 1/3 of the variance accounted for by this solution. Thus, about 2/3 of the scores' variance is found in secondary factors thought to represent specific abilities.

Table 2. The First 13 Unrotated Factors of 33 Cognitive Test Variables: Longitudinal Initial Sample

Var	1	2	3	4	5	6	7	8	9	10	11	12	13	h^2	Label
	FACTOR														
1	.13	-.42	-.04	.03	-.33	.17	.48	.16	.29	-.15	.03	.25	.11	.78	Simple React Time DT
2	-.07	-.03	.06	.00	.15	-.62	.30	-.12	-.31	.40	.35	.12	.22	.98	Simple React Time PC
3	.16	-.52	-.11	-.12	-.29	.20	.46	.12	.11	.20	.14	-.04	-.04	.75	Choice React Time DT
4	-.11	-.08	.20	-.02	.06	-.61	.41	.05	.09	-.12	-.37	-.07	-.45	.98	Choice React Time PC
5	.21	-.67	.21	-.35	.06	.07	-.12	-.09	-.16	.03	-.07	-.20	-.01	.76	Perceptual Spd/Acc DT
6	-.12	-.41	.49	-.41	.09	-.07	-.13	.09	-.06	-.08	-.02	-.25	.00	.70	Perceptual Spd/Acc PC
7	.28	-.57	-.02	-.13	-.06	.20	.20	-.08	-.26	-.11	.17	-.23	-.06	.67	Short-Term Memory DT
8	-.25	-.03	.36	-.22	.07	-.24	-.08	.34	.51	.07	.02	-.15	.37	.86	Short-Term Memory PC
9	.51	-.49	.26	-.19	.14	.12	-.06	-.12	-.07	.13	-.05	.14	-.05	.70	Target Identification DT
10	-.18	-.36	.24	-.39	.06	-.05	-.25	.10	-.01	-.03	-.04	.63	-.14	.87	Target Identification PC
11	.44	-.24	-.21	-.39	-.14	-.13	-.24	.19	-.06	.04	.07	-.01	.02	.58	Number Memory DT
12	-.38	-.09	.47	-.12	.19	.13	.02	-.27	.25	.18	-.05	-.04	.03	.61	Number Memory PC
13	.57	-.25	.24	.49	-.21	-.07	-.17	.03	.03	.12	-.02	.01	-.03	.79	Target Tracking 1
14	.61	-.22	.27	.44	-.16	-.09	-.16	.03	.04	.10	.00	.01	-.03	.76	Target Tracking 2
15	.44	-.22	.10	.42	-.12	-.08	-.14	.10	-.09	.04	.00	.15	.01	.51	Cannon Shoot
16	.36	-.21	.14	.51	-.28	-.06	-.14	.03	-.00	.16	-.20	-.16	.01	.65	Target Shoot
17	.28	-.31	.15	.20	.04	-.25	.05	-.25	-.09	-.65	-.02	.06	.38	.94	Pooled Movement Time
18	-.66	-.02	.10	-.08	-.38	-.12	-.16	.02	.02	-.00	.10	.01	-.02	.65	Assembling Objects
19	-.74	-.12	.08	.07	-.15	-.03	-.07	-.08	.02	-.00	.01	.02	-.00	.60	Map Test
20	-.62	.21	.06	-.18	-.34	-.01	-.02	.10	-.14	-.08	.02	-.05	-.00	.62	Maze Test
21	-.56	.12	.06	-.09	-.44	-.08	-.03	.02	-.21	-.06	-.06	-.02	.02	.59	Object Rotation Test
22	-.65	-.10	.01	-.06	-.30	-.05	-.07	-.18	-.01	.02	.07	-.01	-.02	.57	Orientation Test
23	-.67	.00	.18	.04	-.28	-.07	-.10	.10	.06	-.05	.18	-.03	-.11	.64	Reasoning Test
24	-.70	-.18	.22	.22	.06	.12	.08	-.30	.05	.07	-.02	.02	.01	.75	ASVAB Arith. Reasoning
25	-.56	-.32	-.38	-.01	.00	.01	.01	.03	-.13	.17	-.41	.00	.21	.82	ASVAB Auto/Shop
26	-.20	.29	.53	.07	-.01	.16	.15	.44	-.35	-.03	-.05	.03	.07	.78	ASVAB Coding Speed
27	-.62	-.34	-.29	.15	.12	.01	-.01	.11	-.10	.11	-.26	.04	.18	.75	ASVAB Electronics Inf
28	-.66	-.33	-.23	.28	.24	.01	-.04	.15	.00	-.06	.07	.01	-.06	.76	ASVAB General Science
29	-.77	-.22	-.19	.03	-.03	.02	-.00	-.05	-.04	.06	-.15	.02	.08	.72	ASVAB Mechanical Comp
30	-.65	-.10	.29	.30	.10	.14	.04	-.23	.08	.02	.11	.05	-.06	.71	ASVAB Math Knowledge
31	-.15	.36	.62	.11	.09	.29	.23	.08	-.21	.04	-.13	.06	.08	.77	ASVAB Number Operations
32	-.52	-.25	.05	.29	.27	.02	-.05	.32	-.03	-.09	.23	-.09	-.12	.69	ASVAB Paragraph Comp
33	-.58	-.36	-.17	.29	.26	-.03	-.07	.24	.01	-.10	.17	-.04	-.12	.76	ASVAB Word Knowledge
Eigen-value	7.82	3.03	2.25	2.10	1.36	1.24	1.13	1.00	.92	.87	.79	.75	.71		

The tests loading highest on the first factor are the Map Test (.74), Arithmetic Reasoning (.70), and Mechanical Comprehension (.77). Four of the other five spatial ability tests have loadings greater than .60 on the first factor, and the fifth (Orientation) loads .56. Clearly, there is a major spatial ability influence on the "g" factor defined by this battery of test scores. Three ASVAB subtests load in the .60's: Electronic Information, General Science, and Mathematics Knowledge. An additional three subtests load in the .50's: Auto/Shop, Paragraph Comprehension, and Word Knowledge. Only two ASVAB subtests do not load highly on the first factor: Coding Speed (.20) and Number Operations (.15).

Only seven of the seventeen scores from the computer-administered tests load as high as .35 on the first factor and no loading is greater than .61 (for Target Tracking 2). However, the scores that do load moderately highly come from the psychomotor tests, the Target Identification test, and the Number Memory test. For the Target Identification and psychomotor scores, this may be partially due to the relatively greater amount of spatial ability in the first factor than is often found in analyses of other batteries. We think that spatial ability contributes substantially to psychomotor performance on these tests, and the Target Identification test uses abstracted, rotated drawings of military vehicles and aircraft as its stimuli. Such figures would appear to draw on spatial ability. Another possible explanation is that general learning ability is important for early performance on these tests (generally speaking, no more than forty items are administered on these tests). If this were true, then scores obtained from examinees after they had completed many more items on these tests should have reduced loadings on the first factor in this battery. The moderately high loadings by Number Memory test scores probably reflect the quantitative component of "g" for this battery. In addition, Number Memory is similar to tests of working memory capacity which has been considered as a measure of "g" (Kyllonen & Christal, 1990).

The simpler, computer-administered perceptual tests, including the movement time score, had no scores with loadings greater than .23 on the first factor. Thus, they do not appear to have a high "g" component, at least as measured by this battery.

Table 3 shows the loadings of the ASVAB subtests on the first factor of the solution obtained in this study (called the "Project A Solution" in the table) and for the first factor of an analysis reported in Ree, Earles, & Teachout (1992) which used only the ASVAB subtests on the ASVAB normative sample. They found that the general factor accounts for about 60% of the total variance in the factor solution compared to about 33% of the variance in our solution. We compared the loadings on psychometric "g" in order to explore the effect on ASVAB subtest loadings on "g" when the battery of tests is extended considerably beyond the ASVAB subtests. Note first that the loadings are generally lower for the subtests in this study; the average loading is .54 for the Project A solution and .79 for the normative sample. In addition, the Coding Speed and Number Operations subtests show dramatically lower loadings, .43 and .57 lower, respectively. In contrast, however, the Mechanical Comprehension subtest loading was only two points lower (.77 versus .79).

The generally lower ASVAB subtest loadings could be due to the more complex nature of the first factor in this solution, particularly the addition of the spatial ability component. The much lower CS and NO loadings in this study could be due to the fact that they received relatively high loadings on several of the specific factors, notably factors 3, 8, and 9 for CS and factors 2 and 3 for NO (see Table 2). Although it is risky to interpret unrotated factors, these loadings indicate that CS and NO appear to load on factors defined by the percent correct scores for the computer-administered tests (e.g., Perceptual Speed/Accuracy, Memory, Number Memory).

In conclusion, these findings show that the inclusion of spatial, perceptual and psychomotor test scores with ASVAB subtest scores changes the nature of the first factor, often thought of as a measure of "g", when the scores are subjected to principal components analysis. Although this is not surprising, the nature of the change is informative. Scores from relatively simple, computer-administered measures of perception show small "g" loadings, whereas scores from psychomotor, target identification and number memory measures show moderately high loadings on "g". Scores from spatial tests showed substantial loadings on the first factor due no doubt to the fact that we included scores from six spatial tests--far more spatial tests than are typically administered in multi-aptitude batteries. The pattern of loadings by ASVAB subtests is also altered

when the "g" factor is defined as it was here in comparison to the definition of "g" obtained from analysis of the ASVAB subtests alone. In short, psychometric "g" is a function of the content of the test battery. Some "specific" factor variance, if shared by several tests in the battery (as spatial ability was in our battery) may be captured in psychometric "g".

We have offered some interpretations and speculations about these findings. We plan to conduct hierarchical factor analysis of these data to examine more precisely the nature of the specific factors that appear to make up about 2/3 of the stable variance in the scores.

Table 3. Comparison of First Factor Loadings for ASVAB Subtests

ASVAB Subtest	Project A Solution	ASVAB Normative Sample
Arithmetic Reasoning	70	87
Auto/Shop	56	69
Coding Speed	20	63
Electronic Information	62	82
General Science	66	88
Mechanical Comprehension	77	79
Mathematics Knowledge	65	82
Number Operations	15	72
Paragraph Comprehension	52	81
Word Knowledge	58	87

References

Allen, S. J., & Hubbard, R. (1986). Notes and commentary: Regression equations for the latent roots of random data correlation matrices with unities on the diagonal. Multivariate Behavioral Research, 21, 393-398.

Brogden, H. E. (1951). Increased efficiency of selection resulting from replacement of a single predictor with several differential predictors. Educational and Psychological Measurement, 11, 173-195.

Campbell, J. P., & Harris, J. Introduction. In J. P. Campbell and L. M. Zook (Eds.), Building and retaining the career force: New procedures for accessing and assigning Army enlisted personnel, Annual Report, 1990 Fiscal Year (ARI FR-PRD-90-6). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Jensen, A. R. (1980). Bias in mental testing. New York: The Free Press.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning is (little more than) working memory capacity. Intelligence, 14, 389-433.

Peterson, N. G., Hough, L. M., Dunnette, M. D., Rosse, R. L., Houston, J. S., & Toquam, J. L. (1990). Project A: Specification of the predictor domain and development of new selection/classification tests. Personnel Psychology, 43, 247-276.

Peterson, N. G., Russell, T. L., Hallam, G., Hough, L. M., Owens-Kurtz, C., Gialluca, K., & Kerwin, K. (1990). Analysis of the experimental predictor battery: LV Sample. In J. P. Campbell and L. M. Zook (Eds.), Building and retaining the career force: New procedures for accessing and assigning Army enlisted personnel, Annual Report, 1990 Fiscal Year (ARI FR-PRD-90-6). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Ree, M. J., & Earles, J. A. (1990). The differential validity of a differential aptitude test. (AFHRL-TR-89-59). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Ree, M. J., & Earles, J. A. (1991a). The stability of g across different methods of estimation. Intelligence, 15, 271-278.

Ree, M. J., & Earles, J. A. (1991b). Predicting training success: Not much more than *g*. Personnel Psychology, 44, 321-332.

Ree, M. J., Earles, J. A., & Teachout, M. S. (1992). General cognitive ability predicts job performance. (AL-TP-1991-0057). Brooks AFB, TX: Manpower and Personnel Research Division, Armstrong Laboratory.

Spearman, C. (1927). The abilities of man. New York: MacMillan Co.

Thurstone, L. L. (1938). Primary mental abilities. Chicago: University of Chicago Press.

The Factor Structure of a Spatial Test Battery[1]

Teresa L. Russell
Human Resources Research Organization

Lloyd Humphreys
University of Illinois at Champaign-Urbana

Rodney Rosse and Norman G. Peterson
American Institutes for Research

Six paper-and-pencil spatial tests were developed in the Army's Project A: Assembling Objects, the Map Test, the Maze Test, Object Rotation, Orientation Test, and the Reasoning Test (Peterson, Hough et al., 1990). In a follow-on longitudinal validation project, the Career Forces project, these tests were administered to more than 40,000 new Army recruits. Three of the six spatial measures are now a part of the Enhanced Computer Assisted Test (ECAT) battery of tests that the Services are considering for inclusion in the Armed Services Vocational Aptitude Battery (ASVAB). These three tests are: ECAT Assembling Objects, ECAT Orientation, and ECAT Figural Reasoning. This paper explores the factor structure of the six spatial tests. Exploratory, confirmatory, and second-order factor analyses were conducted on Career Force sub-samples of 4,000 or more recruits. Exploratory analyses yielded one or at most two spatial factors. Confirmatory analysis provided support for up to three spatial factors. A Schmid-Leiman, second-order analysis yielded a strong general factor and two specific factors, with four tests loading on them modestly.

Researchers first identified a spatial factor, distinct from verbal ability, during the 1920s and 1930s. This factor underlying spatial tests (e.g., pattern perception, mazes) was called perceptual ability (Brown & Stephenson, 1933), practical ability, or simply "k" (Smith [1934] reported in Smith, 1948). "Space" was a label applied by Thurstone (1938). He administered 56 tests, designed to tap a wide range of abilities, to 218 subjects. He extracted 13 factors but could only label nine: Perceptual Speed, Number, Verbal Relations, Word Fluency, Memory, Induction, Reasoning, Deduction, and Space. Five tests with the highest loadings on the Space factor were Flags, Lozenges B, Cubes, Pursuit, and Surface Development--all of which require the ability to imagine the transformation of an object or figure in space.

In the 50 years since Thurstone's initial work, most spatial abilities research has focused on defining the number and structure of spatial subabilities rather than the existence of a broad spatial construct. Numerous studies have yielded at least one spatial factor. Three spatial factors have strong support--Visualization, Spatial Orientation, and Speeded Rotation--and several other factors have some support (Ekstrom, French & Harman, 1979; Guilford & Lacey, 1947; Lohman, 1979, 1988; McGee, 1979).[2]

---

[1]This research was funded by the U.S. Army Research Institute for the Behavioral and Social Sciences, Contract Nos. MDA903-82-C-0531 and MDA903-89-C-0202. All statements expressed in this paper are those of the authors and do not necessarily reflect the official opinions or policies of the U.S. Army Research Institute or the Department of the Army.

[2]Unfortunately, authors have not labelled factors consistently. For example, McGee (1979) refers to the factor defined by Thurstone's Flags, Figures, and Cards as Visualization, whereas Guilford and Lacey (1947) named it Spatial Relations. Lohman (1988) refers to it as Speeded Rotation and others (Ekstrom et al., 1979) have used the name Spatial Orientation. It is, therefore, very important to consider the marker tests as well as the label and definition researchers apply in defining factors. Lohman (1988) appears to have used labels that are most true to prior research efforts. In this report, his labels are used for groups of marker tests that tend to load together on factors.

Visualization, Vz, is the "ability to manipulate or transform the image of spatial patterns into other visual arrangements" (Ekstrom et al., 1979, p. 41). Visualization underlies complex spatial tasks that are relatively unspeeded, such as paper-folding, paper form board, surface development, block design, mechanical principles, and three-dimensional rotation tests. In this framework, ECAT Assembling Objects is a Visualization test. Assembling Objects has two types of items; both types require the subject to figure out what an object will look like when its parts are put together. Half of the items are form board items, like puzzle pieces; the other half are geometric figures (e.g., squares, circles) that must be assembled in a specific way. Lohman (1988) and others (Guilford & Lacey, 1947) have noted that figural reasoning tests often load on this factor.

Spatial Orientation (SO) involves reorienting an imagined self; that is, "subjects must imagine they are reoriented in space and then make some judgment about the situation" (Lohman, 1979, p. 188). Marker tests for Spatial Orientation include Aerial Orientation (Guilford & Lacey, 1947), ECAT Orientation, and the Project A Map tests (Peterson, Hough et al., 1990). For example, each item on the Aerial Orientation test shows a cockpit view of a shoreline. Pictures of an airplane at different altitudes are also presented. Subjects must identify the picture of the airplane that would produce the cockpit view provided. In the Map test, subjects are given a map. With each new item the subject is dropped to a new location on the map and instructed to reach a specific objective. Subjects must indicate the appropriate direction (e.g., NW, SW) to reach the objective. The ECAT Orientation test involves reorienting a picture to match a frame. This task, and most other orientation tasks like it, can be accomplished by mentally rotating parts of the object, rather than reorienting oneself. Lohman (1979, 1988) suggests that most orientation tests can also be solved with a rotation strategy.

Speeded Rotation (SR) is defined by tests such as Flags, Figures, and Cards (Thurstone & Thurstone, 1941) that involve rapidly rotating a stimulus (in the picture plane). The Project A Object Rotation test is a test of Speeded Rotation. More difficult rotation tests, involving three dimensions or rotation in the depth plane, often load with more complex tests on Visualization (Lohman, 1988). Speeded Rotation, sometimes called Spatial Relations, is probably one of the most consistently and cleanly identified spatial factors; it emerges in virtually all studies where "two dimensional" rotation tests are used. Compared to Visualization and Spatial Orientation, it is a narrow factor measuring a fairly specific ability.

At least three other spatial constructs have some factor-analytic support: Flexibility of Closure (Cf), Speed of Closure (Cs) and Spatial Scanning (Ss) (Ekstrom et al., 1979; Lohman, 1988). Flexibility of Closure involves breaking one gestalt to form another (to locate concealed figures in a distracting environment, for example). Hidden Patterns published by the Educational Testing Service (ETS) is a marker test. Speed of Closure, which sometimes combines with Cf in factor solutions, requires the ability to unify an apparently disparate perceptual field into a single percept (Ekstrom et al., 1979); ETS's Gestalt Completion is an example marker test. Spatial Scanning is marked by maze-tracing or path-finding tests and involves the ability to find an appropriate path (Ekstrom et al., 1979; Lohman, 1988). The Project A Maze Test is, for example, a Spatial Scanning test (Peterson, Hough et al., 1990).

Methods and Results

Six paper-and-pencil spatial tests were developed in the Army's Project A: Assembling Objects, the Map Test, the Maze Test, Object Rotation, Orientation Test, and the Reasoning Test (Peterson, Hough et al., 1990). In a follow-on longitudinal validation project, the Career Forces project, these tests were administered to more than 40,000 new Army recruits (Peterson, Russell et al., 1990). Table 1 provides basic psychometric information about the tests.

The goal of this analysis was to investigate the factor structure of the spatial test battery. Particularly, we wanted to know (a) whether it would be useful to group the tests into three factors for scoring purposes and relatedly, (b) the magnitude of the contribution of specific factors over and above that afforded by a general spatial factor. If specific factors account for little of the variance in the solution, multiple test score composites would be unnecessary; tests could be pooled to form one composite.

We examined the factor structure of the spatial test battery in two steps. First we investigated three "models" of the first-order factor structure of the spatial battery, using LISREL (Joreskog & Sorbom, 1986) to compare the models. Then we applied the Schmid-Leiman (1957) procedure to investigate the second-order factor structure.

Investigation of First-Order Factor Models

Three first-order factor models were compared. Model 1 had one factor formed by all six spatial tests. Model 2 included two factors: (1) Speed (composed of the Maze and Object Rotation tests) and (2) Power (including all four of the other spatial tests). Model 2 was of interest because previous exploratory factor analyses had yielded those two factors (Peterson, Russell et al., 1990). In Model 3 the Project A tests were organized according to the three major spatial factors identified in spatial research literature. It had three factors: (1) Speed (composed of the Maze and Object Rotation tests), (2) Visualization (including Assembling Objects and Figural Reasoning), and (3) Orientation (subsuming the Orientation Test and the Map Test).

The LISREL analyses, as shown in Table 2, suggested that the second or third models might be useful ways to summarize spatial test scores. Both models show improved fit over the single-factor model. We expected Model 2 (the speed/power distinction) to fit well because it is the model suggested by exploratory analyses. For Model 3, the chi-square value was reduced substantially (from 235.62 to 19.62) with a loss of three degrees of freedom. Also, note that for Model 3, the correlations between the factors (see the Phi matrix) are not extremely high and suggest that the three factors may measure somewhat different constructs. Because the three-factor solution was conceptually appealing and appeared to be psychometrically meaningful, we opted to carry three first order factors into the second-order analysis.

Higher-Order Analysis

We used the Schmid-Leiman transformation to place both first order and second order factors in a single order consisting of a general factor and orthogonal group factors. The results appear in Table 3. As shown, all tests

have large loadings on the second-order general factor. Loadings on the Speed and Orientation specific factors are modest, and loadings on the Visualization factor are essentially zero, suggesting that virtually all reliable variance in the Assembling Objects and Reasoning tests is tapped by the general factor.

Discussion

Although there was some support for the three major factors identified in previous research, most of the data pointed toward forming one factor (for all tests). All six tests have strong loadings on the general factor (ranging from .62 for Maze to .75 for Assembling Objects), and generally, broad factors are likely to be better predictors than narrow factors. Also, including more than one spatial composite in prediction equations will reduce degrees of freedom, a consideration that may be important for analyses where Ns are small. Finally, if it were necessary to use only one spatial test in an operational predictor battery, the Assembling Objects and Reasoning tests, which are good measures of the general factor, should be primary candidates.

References

Brown, W., & Stephenson, W. A. (1933). A test of the theory of two factors. British Journal of Psychology, 23, 352-370.

Ekstrom, R. B., French, J. W., & Harman, H. H. (1979). Cognitive factors: Their identification and replication. Multivariate Behavioral Research Monographs, 79, 1-84.

Guilford, J. P., & Lacey, J. I. (1947). Printed Classification Tests, A.A.F., aviation psychological progress research report, 5. Washington, DC: U.S. Government Printing Office.

Joreskog, K.G., & Sorbom, D. (1986). LISREL VI: Analysis of linear structural relationships by maximum likelihood, instrumental variables, and least squares methods. Morresville, IN: Scientific Software.

Lohman, D. F. (1979). Spatial ability: A review and reanalysis of the correlational literature (Aptitude Research Project Technical Report No. 8). Stanford, CA: Stanford University.

Lohman, D. F. (1988). Spatial abilities as traits, processes, and knowledge. In R. J. Sternberg (Ed.), Advances in the psychology of human intelligence (Vol. 4, pp. 181-248). Hillsdale, NJ: Lawrence Erlbaum Associates.

McGee, M. G. (1979). Human spatial abilities: Psychometric studies and environmental, genetic, hormonal, and neurological influences. Psychological Bulletin, 86, 889-918.

Peterson, N. G., Hough, L. M., Dunnette, M. D., Rosse, R. L., Houston, J. S., & Toquam, J. L. (1990). Project A: Specification of the predictor domain and development of new selection/classification tests. Personnel Psychology, 43, 247-276.

Peterson, N. G., Russell, T. L., Hallam, G., Hough, L. M., Owens-Kurtz, C., Gialluca, K., & Kerwin, K. (1990). Analysis of the experimental predictor battery: LV Sample. In J. P. Campbell and L. M. Zook (Eds.), Building and retaining the career force: New procedures for accessing and assigning Army enlisted personnel, Annual Report, 1990 Fiscal Year (ARI FR-PRD-90-6). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Schmid, J., & Leiman, J.A. (1957). The development of hierarchical factor solutions. Psychometrica, 22, 1, 53-61.

Smith, I. M. (1948). Measuring spatial abilities in school pupils. Occupational Psychology, 22, 150-159.

Thurstone, L. L. (1938). Primary mental abilities. Chicago: University of Chicago Press.

Thurstone, L. L., & Thurstone, T. G. (1941). The Primary Mental Abilities Tests. Chicago: Science Research Associates.

Table 1
Psychometric Properties of Paper-and-Pencil Spatial Tests

Test[1]	No. of Items	Number Correct: Mean[2]	SD	Internal Consistency Alpha: PTB[3]	TB[4]	EB[5]	Test-Retest: PTB[6]	TB[7]
Assembling Objects	36	23.55	7.15	.92	.90	.88	.74	.70
Object Rotation	90	59.13	20.15	.97	.97	.98	.75	.72
Maze	24	16.95	4.85	.89	.89	.90	.71	.70
Orientation	24	12.25	6.21	.88	.89	.89	.80	.70
Map	20	7.86	5.45	.90	.89	.88	.84	.78
Reasoning	30	19.53	5.44	.83	.86	.85	.64	.65

Table 2
LISREL Runs on Initial Longitudinal Sample Spatial Test Data to Examine Three First-Order Factor Models

Input Correlation Matrix (N = 4723 Army recruits)

Test	Assembling Objects	Map	Maze	Object Rotation	Orientation	Reasoning
Assembling Objects	1.00					
Map	.51	1.00				
Maze	.48	.40	1.00			
Object Rotation	.44	.40	.50	1.00		
Orientation	.49	.52	.39	.40	1.00	
Reasoning	.55	.50	.45	.41	.47	1.00

Definitions of Alternate First-Order Factor Models

(1) One first-order factor of all spatial tests.
(2) Two first-order factors: a) Speed (Maze Test, Object Rotation Test) and b) Power (all other tests).
(3) Three first-order factors: a) Speed (Maze Test, Object Rotation Test), b) Visualization (Assembling Objects, Reasoning Test), c) Orientation (Orientation Test, Map Test).

Generalized Least Squares LISREL Results

Model	Coefficient of Determination	df	Chi Square	Goodness of Fit	Adjusted Goodness of Fit	Root Mean Square Residual
1	.869	9	235.62	.983	.961	.033
2	.902	8	79.57	.994	.985	.017
3	.917	6	19.62	.999	.995	.008

[1]Object Rotation and Maze Tests are designed to be speeded tests. Alpha is not an appropriate reliability coefficient but is reported here for consistency. Correlations between separately timed halves for the Pilot Trial Battery were .75 for Object Rotation and .64 for Maze (unadjusted).
[2]Initial subsample of the longitudinal subsample; N = 6941-6950 Army recruits.
[3]Pilot Trial Battery, N=290.
[4]Trial Battery, Concurrent Sample, N= 9332-9345.
[5]Experimental Battery, longitudinal sample, subsample N = 6754-6950.
[6]N = 97-125.
[7]N = 499-502.

Table 2 Continued

LISREL Runs on Initial Longitudinal Sample Spatial Test Data to Examine Three First-Order Factor Models

Phi Matrixes for Models 2 and 3: Estimates of True Score Correlations

Model 2

	Power	Speed
Power	1.00	
Speed	.85	1.00

Model 3

	Speed	Visualization	Orientation
Speed	1.00		
Visualization	.86	1.00	
Orientation	.78	.92	1.00

Table 3
Second-Order Analysis of Spatial Test Scores: Schmid-Leiman Transformation

Loadings on the three oblique first-order factors

Test	Speed	Visualization	Orientation
Assembling Objects	.000	.756	.000
Map	.000	.000	.739
Maze	.724	.000	.000
Object Rotation	.686	.000	.000
Orientation	.000	.000	.708
Reasoning	.000	.723	.000

Loadings for first-order factors on the second-order factor

Speed	.862
Visualization	.996
Orientation	.927

Results

	General	Specific Factors		
Test	Factor	Speed	Visualization	Orientation
Assembling Objects	.753	.000	.065	.000
Map	.685	.000	.000	.278
Maze	.624	.367	.000	.000
Object Rotation	.592	.347	.000	.000
Orientation	.656	.000	.000	.266
Reasoning	.720	.000	.062	.000

GENERAL APTITUDE AND SPECIFIC LANGUAGE LEARNING APTITUDE

John Thain
Defense Language Institute

Background

The Defense Language Institute Foreign Language Center (DLIFLC) is the proponent for the current Defense Language Aptitude Battery (DLAB); it is also the agency responsible for basic language training within DoD. DLAB is used to screen candidates for language training who have previously met ASVAB (Armed Services Vocational Aptitude Battery) requirements for language-related occupational specialities.

Apart from completion of training, the primary criteria for success in language training are ratings on the Defense Language Proficiency Tests (DLPTs), a series of measures of listening, reading, and speaking skills. Testing materials are available in forty languages.

This paper reports on two cooperative efforts by DLIFLC and ARI (Army Research Institute) to evaluate the combined use of ASVAB and DLAB as predictors of language training success.

Study I: The Language Skill Change Project

In 1986, DLIFLC and ARI undertook a broad joint research project called the Language Skill Change Project(LSCP), one objective of which was to identify optimal predictors of success in language training. Predictors other than ASVAB and DLAB were investigated. The population sample in this study included 1900 Army students in DLIFLC Spanish, German, Russian, and Korean basic language courses. Each of these four languages represent one of the four language difficulty categories at DLIFLC.

As indicated in Table 1, the higher the category of difficulty to which a language belongs, the higher the minimum DLAB cut score for that language. In the more difficult language categories, more training time is allotted to achieve corresponding functional language proficiency. In addition, the more difficult a language, the more its writing system, sound system, and vocabulary tend to differ from those of the English language.

TABLE 1
LANGUAGE CATEGORIES

Language Category	DLAB Cutoff	Representative LSCP Language	Weeks Training
I	85	Spanish	26
II	90	German	34
III	95	Russian	47
IV	100	Korean	47

The overall design of the LSCP is shown at Figure 1.

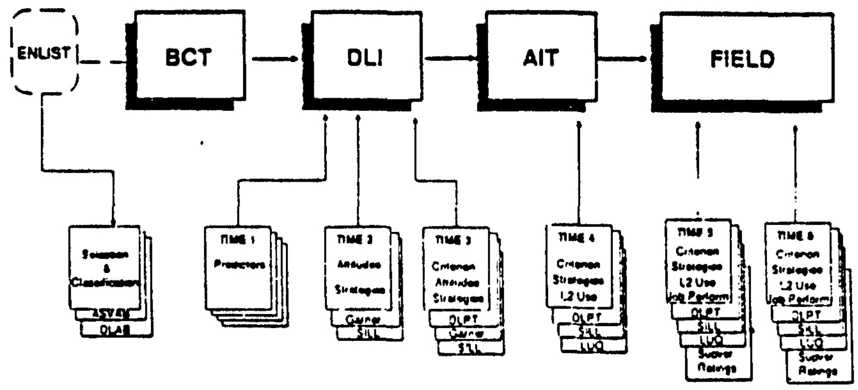


Figure 1

Some of the predictors were administered at enlistment, and others at Time 1 or Time 2. The criterion measures (the DLPTs in each language) were administered at Time 3, after completion of DLIFLC language training. Multiple regression analyses were conducted to determine the incremental contributions of the predictors. Careful consideration was given to the order in which variables were entered into the regression analysis.

Measures already being used to select and place DLIFLC students (ASVAB followed by DLAB) were entered first. The next predictors entered were demographic measures for which data were already available--sex, education, and age. Next to be included were additional demographic data that could be obtained at low cost--prior foreign language training, prior foreign language proficiency, and handedness. Finally variables were entered that required administration and scoring of additional instruments; this included instruments administered at Time 1 before training and instruments administered at Time 2 during training. These additional variables included both measures of stable student characteristics and also measures of motivation and behavior related to a particular point in training.

In all, 16 equations were generated. There were four equations for each of the four languages--one equation for each of the four criteria in each language--attrition, and listening, reading, and speaking skills.

The results of the analyses may be summarized as follows:

(1) Listening and reading skills were better predicted than were speaking skills and the incidence of student attrition;

(2) ASVAB GT and DLAB complemented each other in predicting criterion performance; ASVAB GT furnished higher initial prediction in the easier Spanish and German courses, while DLAB incremental prediction compensated for low ASVAB GT prediction in Russian and Korean;

(3) Additional demographic information concerning previous foreign language study and initial foreign proficiency also contributed to prediction;

(4) Other measures administered during language training--measures of motivation to learn and attitude toward learning, and measures of language learning strategy use--also contributed significantly to criterion prediction. While not directly bearing on selection policy, this last finding has served as a catalyst for a learning strategies project at DLIFLC.

The following tables present an overview of the results, which although simpflified, nevertheless captures these main points.

TABLE 2
MEDIAN R^2 INCREMENT FOR EACH PREDICTOR BLOCK ACROSS 4 LANGUAGES

Order of Entry	Predictor	Reading	Listening	Speaking	Attrition
1	ASVAB GT	.08	.08	.00	.02
2	DLAB	.10	.06	.03	.03
3	Demographics	.01	.01	.01	.04
4	New Demographics	.05	.08	.05	.02
5	Other Variables	.14	.14	.11	.07
	MEDIAN R^2	**.37**	**.34**	**.24**	**.19**

TABLE 3
RANGE OF INCREMENTAL R^2 CONTRIBUTION FOR ASVAB GT AND DLAB ACROSS PROFICIENCY CRITERIA

Order of Entry	Predictor	Spanish	German	Russian	Korean
1	ASVAB GT	.00-.16	.02-.16	.00-.06	.00-.02
2	DLAB	.02-.11	.02-.09	.01-.08	.01-.12
	N	***160***	***182***	***371***	***233***

(All multiple R^2 values for predicting attrition also fell within these ranges, but with greater N because of inclusion of attrition cases.)

Study 2: Compensatory model using ASVAB and DLAB

Background. As the LSCP was being completed, four parallel developments were taking place. First, ARI was conducting additional studies confirming that ASVAB composites and DLAB serve complementary roles in prediction of language proficiency. Some of these studies were conducted at the request of, and in coordination with, DLIFLC.

Second, at the direction of the General Officers Steering Committee(GOSC), which provides policy guidance for the Defense Foreign Language Program(DFLP), DLIFLC was setting DLPT performance goals for its component language schools in all four language categories. A stated goal for individual students is to achieve a "Level 2 rating" in two of the three skills of listening, reading,and speaking as measured by the DLPT--known as the "2/2 standard." The content of standard level proficiency ratings is determined by the Interagency Language Roundtable (ILR), an organization representing all major government language schools, including those outside DOD. A long-term DLI organizational goal is for 80% of the students studying languages in each language category to meet the "2/2 standard." With somewhat less formality, DLIFLC has also set goals for reducing attrition.

As indicated above, DLIFLC instructors in the more difficult language categories have some advantages; traditionally basic courses in more difficult languages have longer course lengths. In addition, DLIFLC, with the assistance of the GOSC, has with increasing success persuaded the Services to insure that all students assigned to the more difficult languages have the recommended DLAB minimum score for these languages. Furthermore, DLIFLC has projected that language departments teaching the more difficult

languages would need more time to make curricular and instructional changes to achieve the "2/2 goals." It should be noted here that another purpose of the LSCP project, which was discussed earlier, was to identify changes in training that might facilitate achievement of these goals.

Third, DLIFLC has arranged for ARI to regularly analyze data and develop briefing slides which illustrate the complementary role that ASVAB and DLAB could play in predicting achievement of DLIFLC training goals. These slides, which have been prepared by Drs. Len White and Jay Silva, take the form of two-way expectancy tables with intervals of DLAB score on one axis and intervals of AFQT(Armed Forces Qualification Test) percentile scores on the other axis; the individual cells have information related to the proportion of attrition cases or the proportion of cases meeting the 2/2 standard. Tables 4 and 5 present the percentage of cases meeting the 2/2 standard in Category I and Category IV languages during a three year period (1987-1990). In order to portray the trends more clearly, cells with less than 35 cases have not been reported.

TABLE 4
PROPORTION OF 2/2 GRADUATES
CATEGORY I LANGUAGES (1987-1990)
(N=1049)

		DLAB INTERVAL					
		<90	90-94	95-99	100-104	105-109	>109
	<65	.21	*	*	*	*	*
	65-74	.30	.37	*	*	*	*
AFQT	75-84	.40	.49	.53	.47	*	*
	85-92	.56	.49	.65	.65	.67	.73
	93-99	*	.60	*	*	*	.86

** cells with less than 35 cases*

TABLE 5
PROPORTION OF 2/2 GRADUATES
CATEGORY IV LANGUAGES (1987-1990)
(N=1197)

		DLAB INTERVAL					
		<90	90-94	95-99	100-104	105-109	>109
	<65	*	*	*	*	*	*
	65-74	*	*	*	*	*	*
AFQT	75-84	*	*	.08	.06	.16	.19
	85-92	*	*	.15	.16	.16	.30
	93-99	*	*	*	.13	.23	.32

** cells with less than 35 cases*

A comparison of Tables 4 and 5 shows significant differences between the two categories of languages in achievement of the 2/2 goal. In addition, Table 5 shows the relative success of DLIFLC and the GOSC in encouraging the Services to send high DLAB students to Category IV languages.

Compensatory Model. ARI and DLIFLC have agreed that the next logical step would be to develop a computer program that could serve as a job aid to Army PERSCOM in assigning students to different categories of languages. This computer program would use a compensatory model involving both ASVAB and DLAB as predictors to optimize language assignment. (Although AFQT scores were used in the briefing slides, the actual model would probably use another ASVAB composite.) ARI personnel have proposed to create a personnel assignment algorithm using integer linear programming or a network solution. Such a program would provide a language category assignment for every individual in an available cohort submitted for analysis. Recent discussions between DLIFLC and ARI have focused on the nature of the constraints on optimal assignment.

Constraints on Model. At one extreme, one possibility would be to set no constraints either on (1) minimum predictor scores for ASVAB and DLAB (thus obviating the advantage of higher DLAB cut scores for more difficult languages, an advantage which has probably contributed to improved training success in these languages) or on (2) the extent to which higher criterion means might be projected in one language category at the expense of other language categories. At the other extreme, one might prescribe differential predictor cutoffs for ASVAB and DLAB up front, and/or require a solution predicting equal success on the DLPTs across language categories.

As to the former case, DLIFLC may not wish to accept a solution with no constraints on projected criterion means across language categories. An example will illustrate why. Compared to all the other prediction equations across languages and criteria, studies indicate that ASVAB and DLAB have the highest predictive power for listening and reading DLPT scores in Category I and II languages. These findings suggest the possibility that DLIFLC could increase the total number of cases meeting the "2/2 standard" across DLIFLC as a whole by assigning substantially more students with higher ASVAB and DLAB scores to Category I and II languages. Yet such a solution would not necessarily be optimal from the point of view of DLIFLC, if concentrated gains in Category I and II languages were accompanied by little positive or even negative effects on criterion means in other categories.

As to the latter case, it may be unrealistic to set a constraint of equal predicted success on the DLPTs across language categories, given the current differences in DLPT performance across categories. Although optimal assignment could play a role in increasing performance in the more difficult languages, it may not be possible for changes in language assignment alone to bring performance in the more difficult languages to the same level as the Category I languages; it is very probable that significant instructional and curricular changes would also have to come into play. Thus, a practical model may have to be found between these two extremes.

Need for Flexibility. The recent trend has been for DLPT scores to rise. Any procedure for implementing a computerized job aid for language personnel assignment should be flexible enough to allow review and updating as new data on training success become available.

Furthermore, it would be desirable if a revised selection and classification model could be adapted to include the use of measures in addition to the current ASVAB and DLAB tests. For example: (1) Results from the LSCP and from studies conducted by ARI indicate that previous proficiency in the foreign language to be studied (or previous experience with any foreign language) may contribute significant incremental prediction beyond ASVAB or DLAB. (2) DLIFLC has developed a first draft of new materials to supplement or replace the new DLAB; validation administration of these new test materials may furnish additional prediction data. (3) Recent research indicates that measures of working memory are predictors of success in language training. Current validation efforts of the ECAT (Enhanced Computer Administered Tests) working memory tests and the Armstrong Laboratory CAM (Cognitive Abilities Measurement) battery working memory tests may suggest changes in language personnel screening procedures.

If a computerized personnel assignment model works well with Army linguists, DLIFLC hopes that the procedures will be helpful in designing similar systems to serve linguists in other Services.

References

Clark, John L.D., O'Mara, Francis E. (1991). Measurement and Research Implications of Spolsky's Conditions for Second Language Learning. *Applied Language Learning*, 2:1, 71-112.

Gardner, R.C., Lalonde, R.N., Pierson, R. (1983). The sociocultural model of second language acquistion: An investigation using LISREL causal modeling. *Journal of Language and Social Psychology*, 2(1), 1-15.

Keesling, J. Ward, Lett, John A., Thain, J.W. (1992). *Socio-Educational Models of Second Language Learning: Analysis of LSCP Data.* Presentation at the Annual Meeting of the National Council on Measurement in Education.

Lett, John A., O'Mara, Francis E. (1990). *Foreign Language Learning and Retention: Interim Results of the Language Skill Change Project.* PRC Inc. Work performed under contract for Defense Language Institute Foreign Language Center.

Lett, John A., Thain, J.W. (1991). *The Learning Strategies Project Status Report and CY92 Plan* (ESR Report No. 91-02). Defense Language Institute Foreign Language Center.

Rumsey, Michael, Silva, Jay M., White, Leonard A. (1991). *Relationship of Cognitive Aptitudes to Success in Foreign Language Training.* Unpublished manuscript. U.S. Army Research Institute.

Schmidt, F.L., Hunter, J.E., Larson, M. (1988). *General cognitive ability vs. general and specific aptitudes in the prediction of training performance: Some preliminary findings.* Paper prepared for Navy Personnel Research and Development Center (Delivery Order 0053).

Schute, Valerie J. (1992). *Learning Processes and Learning Outcomes* (AL-TP-1992-0015). Brooks Air Force Base, TX: Human Resources Directorate, Manpower and Personnel Research Division.

Spolsky, Bernard (1989). *Conditions of Second Language Learning.* Oxford:Oxford University Press.

Statman, Mary Ann (1992). *Developing Optimal Predictor Equations for Differential Job Assignment and Vocational Counseling.* Paper presented at the 100th Annual Convention of the American Psychological Association.

Thain, J.W. (1992). *DLAB II Prototype Development Status Report and CY92 Plan* (ESR Report 92-02). Defense Language Institute Foreign Language Center.

What's Past is Prologue

Malcolm James Ree

Armstrong Laboratory[1]

Human Resources Directorate
Manpower and Personnel Research Division

The armed services have a long history of research and development of tests and other measures for selection and classification of personnel (DuBois, 1972). In fact, the fundamental statistical equations showing the limits of selection efficiency and classification efficiency were developed by Brogden (1943, 1946) when he was at the Office of the Adjutant General of the Army. Although many mathematical formulations are important, the two most important selection and classification equations were given by Brogden (1946, 1951) as

$$SE = R \text{ and}$$
$$CE = R(1-r)^{1/2}$$

where SE is selection efficiency which can range from zero to one. R is the validity of the predictor or predictors. *CE* is classification efficiency which can also range from zero to one and *R*, again, is the validity of a linear composite formed by *n* variables and *r* is the average correlations of the composites formed from those n variables. Said differently r is the average correlation of the expected scores. Selection efficiency is solely a function of validity. If only one variable were used, no classification efficiency would be available as the correlation of the expected scores would be one.

Selection efficiency is a simple function of validity while classification efficiency is a joint function of validity and correlation of the scores of linear composites of the predictors. Both are limiting factors. A zero *CE* can be achieved two ways; when R=0, or when r=1.00. A *CE* index of 1.0 can be obtained when R=1.0 and r=0.0. Further it should be emphasized that classification efficiency is not a dichotomous variable stepping abruptly from zero to one. Rather, *CE* is a continuum as the equation suggests. Given a specific set of variables and circumstances, a specifiable *CE* outcome will occur.

An important problem is the notion of equality or inequality of the SE and CE. It is clear the metric is not the same. There must be some non-zero value of *SE* for *CE* to be non-zero. Comparisons are not fruitful. For example, should you prefer a CE of .60 above an SE of .70? In general SE must always exceed CE because of the multiplying effect of r. This is inconsistent with the theory of benefits of classification. Perhaps this seeming contradiction

[1]The opinions expressed are those of the author and not necessarily those of the Department of Defense, Department of the Air Force, nor the US Government.

is the reason for the establishment of the criterion of mean predicted performance (MPP). MPP (see for example Zeidner & Johnson, 1989) is the expectation of the distribution of performance after individuals have been assigned to jobs and the benefits of classification theory can be best realized under optimum classification such as via linear optimization.

Our understanding of this MPP index is incomplete due, at least in part, to statistical artifacts and to a lack of sampling error specification. Statistical artifacts are considered first.

Consider the situation where two equally valuable jobs are to be assigned to individuals via optimization. In one of these jobs there is a high correlation (R) with the predictors; in the other a low correlation with the predictors. In this scenario the reason for the lower correlation could be the poor reliability of one of the criteria. Given this circumstance, the predicted scores for the more poorly predicted criterion will cluster near the mean of the criterion distribution while the better predicted criterion will yield prediction further away from the mean. If the predicted scores were highly correlated then individuals with higher predictor scores could tend to be assigned to the better predicted job and people nearer the mean will be assigned to the less well predicted jobs. The assignment becomes partially determined by predictive efficiency. This yields increased *CE* (and possibly *SE*) but it is, in part, an artifact. This problem needs to be researched.

It is too tempting to consider only point estimates such as R and MPP and forget about sampling variance. In an optimal classification study many sources of error variance can be identified. A non-exhaustive list includes subjects, jobs or job families, criterion sampling, criterion unreliability, predictor sampling, predictor unreliability, sampling of regression weights, estimation error of predicted scores, and error variance in the optimization procedures.

Another problem in interpretation of the MPP is the lack of specification of sampling error. In a meta analyses of several validities (Hunter, Schmidt, & Jackson, 1982, p. 13) of the same test for the same job in several locations or for the same test for several jobs in one location it is recognized that sampling variability causes the validity estimates, usually bivariate correlations, to form a distribution rather than take on a single value. This is directly caused by sampling; it is unavoidable. In validity studies this variability has often been mistaken for differential prediction. However, when the expected sampling variance of the validity coefficients is computed and compared to the observed variance of the coefficients it is usually found that sampling variance accounts for the observed variance. The same logic must be applied to our study of MPP both within and across studies. For example, we inspect the MPP for several jobs or job clusters and observe a range of predicted scores from 80 to 93. Further we observe a variance of these values of 25. How much of that variance is due to sampling variation? One%, 10%, 50%, 99%? A good subject for research.

James Earles suggests a study to investigate the sampling variance of optimization procedures. He suggests creating 10 random "job families," computing regressions in the families and then optimally allocating subjects to jobs. As the job families are randomly equivalent and the regression equations are randomly equivalent, the average increase in MPP and the variance of MPP across the "job families" are explicitly due to sampling error. This

study is close to the form of a jackknife or bootstrap estimation procedure. Closer inspection is warranted. Further, the role of measurement error should be investigated by twice classifying subjects who have test-retest data.

Finally, let us define selection and classification efficiency as S&CE = f(CE, SE) and that the functional form remains unspecified. We have some idea of its nature from empirical studies but an analytic proof is to be preferred. Empirical proof is always dependent on the sample and therefore ,potentially unstable. An analytic proof rests on the foundations of mathematical consistency and is always true within the bounds of the assumptions. This too is a good research topic.

Having observed these facts it is time to think about the future of S&C in the USAF and to investigate the validity and correlations of the predictors at hand and those coming to maturity soon. Let us begin with the paper-and-pencil tests; the Armed Services Vocational Aptitude Battery, ASVAB, and Air Force Officer Qualifying Test, AFOQT. Ree and Earles (1991a, 1991b) have shown these two tests to be highly g saturated and have also shown the expected scores from weighted linear composites to be very highly correlated. These findings indicate that the possible *CE* of the ASVAB or the AFOQT would be limited. Preliminary findings of an in-house study confirm this. The maximum obtainable CE for these tests should be studied if for no other reason than as a baseline of operational practice.

A second area which is progressing toward maturity is the group of computerized tests of the Learning Abilities Measurement Project, LAMP, (Kyllonen, in press a; Kyllonen & Christal, 1990) an advanced study of learning ability using modern cognitive psychology and the cognitive components approach. These components include "working memory," "processing speed," or "time sharing." The most compelling evidence about the nature of these components comes from two recent studies (Kranzler & Jensen, 1991; Miller & Vernon, 1991) which show an almost complete overlap between tests of cognitive components and highly *g* loaded paper and pencil tests. (For a theoretical formulation of the structure of LAMP tests see Kyllonen, in press b) While not sufficient to understand all the consequences, these two studies suggest that cognitive components, although interesting theoretical constructs, may be too closely related to *g* and therefore too correlated with themselves and with tests such as the ASVAB and AFOQT to offer much CE unless they offer greater validity than ASVAB like tests. Again, it is not a question of a dichotomy but rather how much.

Psychomotor tests are often thought of as being taxonomically distinct from paper and pencil predictors (see Fleishman & Quaintance, 1984). Recently, Ree and Carretta (1992) showed that psychomotor tests correlated at least moderately with paper and pencil tests and Carretta and Ree (in press) showed that the incremental validity of psychomotor tests was limited for the prediction of pilot training success. In all, I hold more hope for psychomotor tests than for either aptitude tests or cognitive tests.

In the joint Services Enhanced CAT-ASVAB study, ECAT, (Carey, 1992) new predictors validation was partially based on the suggestion that space perceptual tests might be of some value for increasing *CE*. The work by Carey, and the insights of Alderton and Larson (1992) and the results of Project A (McHenry, Hough, Toquam, Ashworth, & Hanson,

1990) suggest otherwise. Additionally, the cognitive tests of ECAT are based on the components approach and share a common future with the LAMP tests.

The most promising predictor under consideration by the Air Force is personality. Recently, two meta-analyses (Mount & Barrick, 1991; Tett, Jackson, & Rothstein, 1991) suggested that personality variables were predictive of job performance across a wide range of jobs. A forthcoming meta-analysis will show that conscientiousness is predictive of almost all jobs and that it is incremental to aptitude for the prediction of job performance. Carretta and Ree (in press) have shown personality to be incremental to aptitude in the prediction of pilot training success. Given incremental validity and lower correlations with most predictors, more possibility of increased CE exists for personality variables. In the same, vein job preferences/interests should be investigated. The VOICE (Alley & Mathews, 1982) is available and recommended for the task.

If these last several paragraphs sounded like the baleful lament of pessimism or the cry of premature defeat, they were not. They are though, careful consideration of the puzzling obstacles to be overcome in research on classification. These obstacles are but faint shadows of the obstacles we can expect to meet in implementation. We would be ill-advised to press-on without consideration of these as research issues. These and other important issues should form the basis for a cooperative plan for S&C under the TAPSTEM[2] structure.

Important studies are now being conducted which can lead to that cooperative plan and support the services in their quest to increase the utility of manpower. The first is a planning study, the second is a sensitivity analysis of optimum classification techniques based on simulation. An applied study is in progress using MPP as the dependent variable from which we expect to gain practical knowledge.

The study "Building a Joint-Services Classification Roadmap" is jointly sponsored by the Air Force and the Army with monitoring participation from the Navy. Through survey and personal contact it seeks to determine what the goals and priority of manpower planners are and to plan a series of studies across several years to accomplish the research to achieve these goals. Many of you here today have participated in this singular effort and you and your successors can expect to benefit from the study. One of the unique products of the Roadmap study will be a computerized procedure which will allow planning and revision of planning as resources shrink and expand. The Roadmap study is being conducted under contract by HumRRO.

A second study is being conducted by the American Institutes for Research. It is a sensitivity analysis of CE measured by MPP by optimal allocation. This is a simulation study which will vary sample sizes, predictor correlations among themselves and with the criterion as well as number of job clusters. Also psychometric qualities of the predictors and criteria, such as reliability, will be manipulated to determine their effect on MPP. The study iterates through multiple "regression based validity-clustering-optimum allocation studies" to

[2] TAPSTEM (Training and Personnel Science and Technology Evaluation Management) is the name of the oversight structure which coordinates the research of the services to decrease duplication and maximize research and development productivity.

determine the expectation and variance of these procedures. This will estimate the joint error variance of the combination of the conditions (i.e. sample sizes, number of job clusters, validity, predictor correlations). It is conceivable that an N-dimensional table or abac might result which would be a guide to joint distributions of sampling errors and a distribution of outcomes.

The applied study which uses MPP as the dependent variable is an effort to develop new aptitude indexes or selection and classification composites for the Air Force enlisted force. In this effort we have clustered about 150 enlisted jobs into from 2 to 7 clusters. Within each level of clustering there are 5 methods of weighting ASVAB tests: regression weights, only positive regression weight (i.e. 9 ASVAB tests only), and three methods of simple or unit weights based on thresholds of minimum validity for the tests. In these last four weighting methods not all of the ASVAB tests necessarily enter into each composite. Using these regression equations or regressions based on equations for the composites, linear optimization has been used to allocate subjects to job families. The results are preliminary but they show an increase in MPP of about .23 criterion standard deviation units above random allocation of subjects.

By way of summary it is appropriate to say that the American military has long taken an active interest in S&C, especially classification. I have reviewed what I believe to be the likely CE increases from the mature (paper-and-pencil) and maturing (LAMP, psychomotor) testing technologies and can point to at least one, personality, which looks promising for increasing MPP. The need for error specification has been prominently noted as a research opportunity. The Roadmap Study, the sensitivity analysis, and the USAF enlisted composite development study will all add to our store of knowledge and be the foundations for advanced research in the future.

References

Alderton, D. L., & Larson, G. E. (1992). *Dimensions of ability: Diminishing returns?* Paper presented at the ARI Selection and Classification Conference, Alexandria, VA.

Alley, W. E., & Mathews, M. M. (1982). The Vocational Career Examination: A description of the instrument and possible applications. *Journal of Psychology, 112*, 169-193.

Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: A meta analysis. *Personnel Psychology, 44*, 1-26.

Brogden, H. E. (1946). On the interpretation of the correlation coefficient as a measure of predictive efficiency. *Journal of Educational Psychology, 37*, 65-76.

Brogden, H. E. (1951). Increased efficiency of selection resulting from replacement of a single predictor with several different predictors. *Journal of Educational Psychology, 11*, 173-196.

Carey, N. (1992). *New predictors of mechanic's job performance: Marine Corps findings.* Paper presented at the annual meeting of the American Psychological Association, Washington, DC.

Carretta, T. R., & Ree, M. J. (in press). *Pilot Candidate Selection Method (PCSM): What*

makes it Work.

Dubois, P. H. (1972). *A history of psychology testing.* Boston, MA: Allyn and Bacon, Inc.

Fleishman, E., & Quaintance, M. K. (1984) . *Taxonomies of human performance: The description of human tasks.* Orlando, FL : Academic Press.

Hunter, J. E., Schmidt, F. L., & Jackson, G. (1982). *Meta analysis: Cumulating research findings across studies.* Beverly Hills CA: Sage.

Kranzler, J. H., & Jensen, A. R. (1991). The nature of psychometric g: Unitary process or a number of independent processes? *Intelligence, 15,* 397-442.

Kyllonen, P. C. (in press a). CAM: A theoretical framework for cognitive ability measurement. In D. Detterman (Ed.), *Theories of intelligence.* Hillsdale, NJ: Erlbaum.

Kyllonen, P. C . (in press b). *Aptitude testing inspired by information processing: A test of the four-source model.*

Kyllonen, P. C., & Christal, R. E. (1990). Cognitive modeling of learning abilities: A status report of LAMP. in R. Dillon & J. W. Perrigino (eds.), Testing : Theoretical and applied issues. NY: Freeman.

McHenry, J. J., Hough, L. M.., Toquam, J. L., Hanson, M. A., & Ashworth, S. (1990) Project A validity results: The relationship between predictor and criterion domains. *Personnel Psychology, 42,* 335-354.

Miller, L. T., & Vernon, P. A. (1991). The general factor in short-term memory, intelligence, and reaction time. *Intelligence, 16,* 5-29.

Ree, M. J., & Carretta, T. R. (1992). The correlation of cognitive and psychomotor tests. AL-TR-1992-xx. Brooks AFB, TX: Armstrong Laboratory, Human Resources Directorate, Manpower and Personnel Division, Brooks

Ree, M. J., & Earles, J. A. (1991a). The stability of convergent estimates of g. Intelligence, 15, 271-278.

Ree, M. J., & Earles, J. A. (1991b). Estimating psychometric g: An application of Wilk's theorem . Paper presented at the annual meeting of the American Psychological Association, San Francisco, CA.

Tett, B. P., Jackson, D. N., & Rothstein, M. R. (1991). Personality measures as predictors of job performance: A meta-analytic review. *Personnel Psychology, 44,* 703-742.

Zeidner, J., & Johnson, C. (1989). *The utility of selection for military and civilian jobs.* IDA Paper P-1997. Alexandria VA: Institute for Defense Analyses.

SELECTION AND CLASSIFICATION FOR THE CAREER FORCE[1]

Michael G. Rumsey
U.S. Army Research Institute

We are now reaching a watershed in our selection and classification research. A major programmatic effort launched ten years ago is nearing completion. The accomplishments of this program have advanced some avenues of research and laid the foundation for several new ones. Concurrently, advances outside the boundaries of this program have opened new opportunities. Finally, the world in which we are conducting and applying our research has changed dramatically, and our research program needs to reflect these changes. These influences intertwine to shape our program of the future. We now proceed to examine each in turn.

Influence Category #1: The Soldier Selection Projects

First we need to examine the historical foundation for our present research program. The period 1980 to 1982 was a turbulent one in the history of Army selection and classification research. During that time, questions about the norming and job relevance of the Joint Service selection and classification test battery, the Armed Services Vocational Aptitude Battery, or ASVAB, led to a directive from the Department of Defense that the Services participate in a joint research program to validate the ASVAB against job performance. Concurrently, concerns by the Army about whether enough qualified soldiers could be found to man the complicated new weapons systems that would emerge during the decade of the 80's led to a questioning of whether new predictor measures might be needed to supplement the ASVAB.

The convergence of these influences resulted in a set of research projects of epic proportions, projects I will refer to here collectively as Soldier Selection. The first of these was known as Project A. The emphasis in Project A was the concurrent validation of ASVAB and new predictor tests against the criterion of first tour job performance. The second project is Building the Career Force. The emphasis in Career Force is the completion of a longitudinal validation of ASVAB and new predictor tests against both first and second tour performance measures. Both efforts involved participation of a contractor consortium consisting of Human Resources Research Organization (HumRRO), American Institutes for Research (AIR), and Personnel Decisions Research Institute (PDRI).

We are now over halfway through the Career Force project. The longitudinal validation has, to a great extent, confirmed the findings obtained when predictors and criteria were administered concurrently. A model of first tour performance, consisting of two "can do" and three "will do" dimensions, was generated during the concurrent validation (Campbell, McHenry, & Wise, 1990). This model was confirmed in the longitudinal validation (Childs, Oppler, & Peterson, 1992). The ASVAB was highly predictive of technical proficiency in both validations. The new predictors developed during Project A were found to provide incremental validity over the ASVAB, particularly for the "will do" dimensions. The new predictors were found to add less when administered in advance than when administered at the same time, but contributed substantially in either case (Oppler & Peterson, 1992).

Career Force will continue to be a central focus of our research program for the next two years. We will now begin to pay particular attention to second tour performance and its relationship to measures administered at earlier stages in the soldier's career. A six-factor model of soldier performance in the second tour has already been developed based on preliminary data. This model resembles the first tour model in many

[1]Presented at the meeting of the Military Testing Association, October, 1992. All statements expressed in this paper are those of the author and do not necessarily reflect the official opinions or policies of the U.S. Army Research Institute or the Department of the Army.

respects, but places greater emphasis on those elements of leadership, such as training and counseling, which are required at the junior NCO level (Campbell & Zook, 1990). Now the Career Force second tour data collection has been completed and this model can be re-examined. Once the dimensions of second tour performance have been confirmed, these can be related to predictor data collected on soldiers at entry, at the end of training, in first tour and in second tour. We will at that point have an unprecedented data base which we can use to make recommendations for improving the Army's selection, classification, reenlistment and promotion procedures.

However, the influence of the Soldier Selection projects transcends their status as identifiable components of our research program. Several of the measures developed in Project A have been so successful as predictors of job performance that new research efforts have evolved focusing on their further development, refinement and evaluation. These measures include the Assessment of Background and Life Experiences, or ABLE, and a number of new psychomotor and spatial measures.

ABLE measures temperament dimensions such as achievement, dependability, and adjustment. It predicts disciplinary problems, leadership ratings, and attrition to a dramatically greater extent than does the ASVAB. For this reason, we are examining its potential usefulness in a variety of selection and classification contexts. For example, it is being measured for its potential as a supplemental placement measure for Ordnance Missile and Munitions jobs. Also, some discussion has been given to using ABLE items as part of a proposed Joint Service screen to reduce attrition. This discussion has not led to implementation, in part due to concerns about the extent to which ABLE items might be undermined by faking.

These concerns have stimulated research to determine the extent to which faking is a problem and to learn what steps can be taken to control it. We do know that people instructed to fake can raise their scores (Hough, et al., 1990; Young, White, & Oppler, 1991), but we do not yet know the extent to which scores people obtain in operational contexts reflect the temperament dimensions we are trying to measure rather than the impression the test-taker wishes to convey.

Concern for control of faking has stimulated interest in the development of biographical, or biodata measures. Biodata items tend to be more objective and verifiable than temperament items, and the use of this approach may reduce the propensity to fake. Fred Mael and Amy Schwartz (1991) have found evidence that supports the validity of biodata for predicting performance of West Point cadets.

Project A psychomotor and spatial measures also offer exciting possibilities for improved performance prediction. The Project A spatial tests, which include a maze and a map test, focus on ability to deal with the orientation, location, configuration, arrangement and shape of objects. Tests of psychomotor ability, such as a one-hand and a two-hand measure of eye-hand coordination, were developed for computerized administration. Batteries composed of these spatial and psychomotor tests have been found to be highly effective in predicting the accuracy of tank and TOW gunners in simulated firing exercises.

Currently, six of the Project A spatial and psychomotor tests are being examined with three Navy tests by a Joint Service ASVAB Review Technical Committee (ART) as possible new components of a revised ASVAB. Many research questions are being considered in this examination. Two areas that we are paying particular attention to are practice and coaching effects. Henry Busciglio and Dale Palmer (1992) have found that practice and coaching can both increase performance on Project A spatial tests, but that, at least for some tests, practice can mitigate the effects of coaching and coaching can mitigate the effects of practice. Busciglio and Jay Silva are in the process of summarizing relevant information on coaching and practice to assist a Joint Service determination on which, if any, new tests are ready to be implemented.

The Army is concerned with Army-specific paradigms for new test implementation as well as Joint Service possibilities. Accordingly, we have initiated a contract effort with HumRRO to examine and evaluate alternative models for selection and classification. This effort, currently being monitored by Peter Legree, will provide information on where and when particular tests might be most effectively administered and how they might best be used to support selection and classification decisions.

The Soldier Selection research program has profoundly transformed the landscape of our new predictor research. It has helped define both what new research is needed and what research is not needed. No longer is it necessary to determine whether personality or temperament dimensions are important predictors of soldier performance. We now know that they are. We also know much better than before what dimensions are most promising for such prediction. We do not need to do this research again. We know about the predictability of psychomotor and spatial measures for performance in a variety of jobs. Future research in these predictor domains can focus on refinement of measures, resolution of problem areas such as faking or coaching, examination of differential prediction across jobs, and exploration of related issues associated with implementation. We have not closed out research in these domains, but the research issues have been narrowed considerably.

Influence Category #2: External Research Advances

We have discussed the Soldier Selection research projects and their role in defining future research directions. Now we consider another important influence on our research program--advances that have occurred in the general research community. To the extent that these advances fit in with the problems and issues we are facing, they deserve our close attention and exploration.

One area in which significant advances have taken place is that of individual differences. Traditional approaches to the measurement of aptitude have been challenged. Developments in such predictor domains as practical intelligence, self-efficacy, and cognitive complexity offer promising opportunities for further exploration. We have recently initiated a new contract effort, with AIR, HumRRO and Management Research Institute, to identify which of these areas appear to offer the greatest potential for predicting future performance. This effort, to be guided by Clint Walker, is known as Expanding the Concept of Quality in Performance, or ECQUIP. Concurrently, we have an in-house research effort by Busciglio, Legree and Ivey King examining the construct of social intelligence. We will be particularly interested in exploring how the new predictor measures relate to performance of senior noncommissioned officers and junior officers, and will plan our validation strategy accordingly.

The Army's current and future research program also incorporates another theoretical approach known as Differential Assignment Theory, which had its genesis outside the Soldier Selection program but which has advanced by conducting analyses on data collected in this program. Differential Assignment Theory is causing us to re-examine our traditional approaches to classification. The research conducted by Zeidner, Johnson, and their associates suggest that the potential performance gains from classification are enormous, but that we are a long way from achieving this potential with our current system. The conclusions emerging from this research are powerful, provocative and controversial, and will require further scrutiny and development before this research can be translated into action.

In May of this year we held an ARI Selection and Classification Conference with the express purpose of helping to develop an agenda for future research. This conference provided an extraordinary opportunity to examine recent theoretical developments applicable to our program. At that conference, several of the approaches to individual differences which will be examined in the ECQUIP program were discussed, as well as the work on Differential Assignment Theory. However, we anticipate that many of the concepts presented at this conference will have impacts that are not yet reflected in our current research plans. One pervasive theme in this conference was the importance of team performance. While prediction and measurement of individual contribution to team performance are tremendously challenging research tasks, the time to meet this challenge may now be close at hand. This is but one of many themes that emerged from the conference, and we are examining all of them for their potential incorporation into future ARI research efforts.

Influence Category #3: Changes in the World Environment

We have now discussed two influences on our research program--the foundation provided by the Soldier Selection projects and opportunities afforded by advances in the research community which are relevant to Army problems. We now turn to a third type of influence--changes in the world situation.

For decades, we have operated in a world where the dominant threat to American national security was represented by a single nation--the Soviet Union. Now there is no Soviet Union, the threat has changed, and all the U.S. military services are downsizing, including the Army. These changes do not alter the importance of the issues we are already considering, but they do raise new ones for our attention. The first concerns the nature and structure of the jobs which the selection and classification system is filling. As we downsize, the number of different Army jobs is declining. As the Army's mission changes, the nature of its jobs is changing as well. We need to develop procedures to help the Army restructure its jobs, and have initiated contract research with Akman Associates and Hay Systems to do so.

It is important to recognize that, while the threat has changed, we have not entered a world in which threat is non-existent. Campaigns in Panama and Kuwait have reminded us, if any reminder was indeed needed, that the ultimate military performance criterion is combat effectiveness. We need to insure that our selection measures are at least as valid for wartime as for peacetime.

The difficulties of validating selection measures against combat performance are obvious. Following Desert Storm, we did manage to obtain performance ratings on a small number of combatants, which we are relating to a variety of predictor measures administered in the Soldier Selection projects. Meg Matyuf is presenting some preliminary results from this research at this conference. For the longer term, we are focusing on such questions as what factors represent combat performance, how they are measured, and how they differ from dimensions of peacetime performance. Recently, Elizabeth Brady travelled to Kuwait with contractors from Continental Systems to collect interview data that will be used to help address these questions.

Clearly, all of these factors--changes in mission, changes in job structures, and changes in the definition of combat performance--have implications for the types of soldiers we select. The problem of selection in the post Cold War era was eloquently addressed by MG Gorden, then Director of Military Personnel Management, at the May ARI Conference on Selection and Classification (1992). General Gorden spoke of the importance of versatility--of needing soldiers who can prepare both for war-fighting and peacekeeping missions. This viewpoint was echoed by others at the conference, resulting in a consensus that flexibility was a characteristic that we ought to examine further. Accordingly, as we continue our efforts to identify and develop important predictors that contribute incremental validity or differential validity across jobs relative to currently available predictors, we will give careful attention to the concept of flexibility.

The changing roles and missions of the Army have tended to enhance the importance of the Special Forces, who contribute much to the flexibility of the Army as an organization. We have just begun a program to work with the Special Forces to address their selection and classification concerns.

While we need to insure that our research program is not out of touch with current world conditions, we need also to make sure we are preparing for a future world which may be different from the present one. For example, the Army is currently taking a relatively small proportion of soldiers from the lowest acceptable category on the existing cognitive screen, the Armed Forces Qualification Test (AFQT), a composite of ASVAB tests. Suppose that circumstances change such that there is a need to take in large numbers of applicants that fall into that category. Would we be able to assess the impact of such a change on soldier performance? Would we be able to recommend optimal placements for soldiers in that category? Research designed to help answer these and related questions is currently being conducted in a contract project with HumRRO, known as Augmented Selection Criteria. The monitor of this effort, Frances Grafton, is also engaged in exploring a number of other ASVAB-related issues, with assistance from Alan Drisko and Bettie Teevan.

Concluding Comments

We have highlighted the major influences on and major components of our research program. The first influence is the Soldier Selection research, which will culminate in the completion of the Career Force project and which has spawned offshoots involving continued research with respect to temperament, biodata, psychomotor and spatial predictors. The second influence is research advances outside ARI, which is reflected in current work on Differential Assignment Theory and new predictors beyond those developed in Project A. The third influence is world events, which impact upon our job restructuring, prediction of combat performance, and Augmented Selection Criteria efforts. Our program is an ambitious one, yet, as the May Selection and Classification conference highlighted, it covers only a fraction of the number of areas that could be profitably pursued. A key element of our future research strategy will be to build on linkages already established with the other research laboratories to ensure that we continue an existing trend toward increased interservice coordination. We are in an era of declining research dollars but not declining research problems, and we need to obtain whatever efficiencies are possible by pooling our efforts and resources to examine common problems through Joint Service cooperation.

References

Busciglio, H. H., & Palmer, D. R. (1992, August). An empirical assessment of coaching and practice effects on three Army tests of spatial aptitude. In Wise, L. L. (Chair), New Predictors of Military Job Performance. Symposium presented at the Annual Convention of the American Psychological Association, Washington, DC.

Campbell, J. P., McHenry, J. J., & Wise, L. L. (1990). Modeling job performance in a population of jobs. Personnel Psychology, 43, 313-333.

Campbell, J. P., & Zook, L. M. (eds.) (1990). Building and retaining the career force: New procedures for accessing and assigning Army enlisted personnel (ARI Technical Report 952). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Childs, R. A., Oppler, S. H., & Peterson, N. G. (1992, August). Confirmatory analysis of two five-factor models of job performance. In Rumsey, M. G. (Chair), Beyond generalizability of small r: Consistency of personnel research. Symposium presented at the Annual Convention of the American Psychological Association, Washington, DC.

Gorden, F. A. (1992, May). Remarks at the ARI Conference on Selection and Classification for the U.S. Army.

Hough, L. M., Eaton, N. K., Dunnette, M. D., Kamp, J. D., & McCloy, R. A. (1990). Criterion-related validities of personality constructs and the effect of response distortion on these validities. Journal of Applied Psychology, 75, 581-595.

Mael, F. A., & Schwartz, A. C. (1991). Capturing temperament constructs with objective biodata. (ARI Technical Report 939). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Oppler, S. H., & Peterson, N. G. (1992, August). Comparing validity results using Project A concurrent and longitudinal samples. In Rumsey, M. G. (Chair), Beyond generalizability of small r: Consistency of personnel research. Symposium presented at the Annual Convention of the American Psychological Association, Washington, DC.

Young, M. C., White, L. A., & Oppler, S. H. (1991, October). Coaching effects on the Assessment of Background and Life Experiences (ABLE). Paper presented at the Annual Conference of the Military Testing Association, San Antonio, TX.

Symposium on Future Selection and Classification Research in the Service Laboratories
<u>FINIS CORONAT OPUS</u>
(MTA 1992)

Dr. Frank Leo Vicino

1. INTRODUCTION

a. <u>Disclaimer</u>

My good friend and colleague, Malcolm, has decided to do me in by billing my presentation as the FINIS CORONAT OPUS

After I thank my generous friend, I must warn you that my answers are not final, will not be wearing a crown and may in fact leave you with more questions before the FINIS CORONAT OPUS in Military Selection and Classification is completed.

b. <u>Any Final System design is made elusive by the ambiguity/changes in forces</u>.

Wally Sinaiko in his paper on Issues and Opportunities for Applied Psychology in an Era of Smaller Force states that the restructuring of the military forces is the most profound change of its kind in forty years.

2. THE SITUATION

These major changes will be affecting the structure of any future military selection and classification system. Lets just look at some of these anticipated changes.

a. <u>New Roles and Missions</u>.

1) General Gorden, the Army Deputy Chief of Staff for Personnel, at the ARI Selection and Classification Conference in May challenged us to help shape the force of the future. He feels that the

future soldier may have to be selected and prepared for very different situations than those faced in the past. The future soldier will be placed in environments like those that faced Gen. Gravalt in LA. and like what is happening now in Yugoslavia. The warrior as peace keeper/war fighter. While we are facing the downsizing we will also have to select, classify and train soldiers and leaders who CAN OPERATE IN A MORE COMPLEX AND RAPIDLY CHANGING ENVIRONMENT. where flexibility and creativity in leadership becomes more important. He ended his presentation with a challenge to Selection and Classification researchers to "PUT SOME MEAT ON THE BONES OF THIS UNCERTAIN SKELETON THAT WE HAVE HANGING IN FRONT OF US."

2) We will have an older and more experienced military force. The downsizing is raising the expertise level and average age of the Navy. In the last three years the average age of the Navy has increased 10%. We have not, however, changed overall manning mixes. Maybe an experienced E5 can do the work of 2 E3's. This increase in expertise could lead to an increase in unit productivity, and we could operate the same number of ships with a downsized more productive force. We might do more with less if we have more brighter more productive personnel. A former Army DCSPER states that we have not recognized the inherent versatility and capability of our young service members-our occupational structures are too narrow For example, the Navy has about 100 ratings and the Army has even more. We need to find ways to expand our human capital. In addition, Sinaiko feels that smaller recruiting goals will result in higher quality military service entrants. At the same time The uncertainties in the military forces has begun to cause some people to leave voluntarily. There is, therefore, a continuing need to preserve the right mix of experience and occupational skills.

3. NEED TO ACCOMMODATE AMBIGUITY/CHANGE

So we do not have the final answer defining the characteristics and composition of this future , possibly constantly changing, force,

What do we know for sure? We know that the incredible shrinking forces will make waves at many levels of selection and classification activities. We may not know to what extent and in what direction, but we do know that there will be changes in the force-mix and it will be necessary that the military S&C process be flexible and quick in its response to these changing needs. Because of this built-in lack of a clear definition of the characteristics of the future warrior and his or her battle system and the fact that these circumstances can change quickly we are left with at least one major research direction. That is, to design and implement a selection and classification system, poised to accommodate rapid changes in expanding personnel needs. Traditional selection and classification testing, like the traditional roles they represent, may need altering.

4. WHAT CAN WE DO WHILE WAITING FOR THE ANSWERS?.

We need a selection and classification system that will accommodate to rapid changes in Force mix and quickly address the need for more expanded and creative personnel measures . Enter TAPSTEM and the computer.

a)TAPSTEM where, by interaction and exposure, our research questions and directions are being defined and clarified. The TAPSTEM opportunities have made the researchers more aware of S&C areas in which we have stable reassuring results and points to areas where questions exist-but are being defined. Excellent programs like the Air Forces ROADMAP and LAMP , the Army's Projects A&B and the Marines Job Performance research are not only providing excellent research answers but more importantly by posing significant research questions are helping to define the future of S&C research. Navy's new program SYMONAC and a proposed program, STAR, hope to integrate findings from TAPSTEM and other research

programs into a Computerized Testing and Assignment of Recruits System.

b)And now computers. The computer-based testing system with its capability for testing new dimensions, introducing new and versatile combinations of scoring paradigms and to accommodate quickly to changing testing needs.enters as an excellent platform for future testing.

The nerve center of this Computerized System is the Computerized Testing Platform that we call CAT which is now operationally used in a number of sites.

A brief outline about the CAT:

1). The Department of Defense in a Joint-Service Program with the Department of Navy as Executive Agent and NPRDC as lead laboratory has developed and implemented a Computerized Adaptive (CAT-ASVAB) version of the conventionally administered Paper and Pencil Armed Services Vocational Aptitude Battery (P&P-ASVAB) that is currently used to select and classify all military service applicants.

2). The major difference between the CAT-ASVAB and the conventional ASVAB rests in a) the way the test is administered and b)the way the items are selected and scored. The items are administered by computer and selected on the basis of the test-taker's response to previous items with the scoring based on the difficulty level of the items answered correctly.

3) An illustrative Example of a Five Item CAT

[Figure here]

4) There are many advantages to this computerized system:

(a) Test Administration (Faster Testing Time, Flexible Testing Sessions, Improved Standardization)

(b) Scoring (Clerical Error Reduction, Immediate Results)

(c) Precision (Better at Extremes of the Ability Range)

(d) Security (no Test booklets to lose, Items in RAM-disappear when computer disconnected, Many items to remember) . But more importantly; tied to the current and pressing need for testing new dimensions, and to accommodate quickly to changing testing needs:

(e) Future Tests- The computer as a test platform offers the opportunity for new and effective testing situations, ,such as presenting test items in which the examinee interacts dynamically with the display- We now have such tests in the form of target-tracking, eye-hand coordination etc. We need to examine however new situational tests of creativity and leadership-characteristics that have been successfully used in an assessment center setting-but which can be much more economically imbedded in a Computer-Administered Mode (Interactive CD-ROM -examples) The computer can take advantage of scoring dimensions such as reaction time etc, learning to learn dimensions, complex composite responses, trends etc.

(f) Simplified Revision-Tests can be revised electronically. Proposed measures can be imbedded in the test battery to supply the test developers with new item data, while not being included in the applicants score. The item data can be used to determine the suitability of the new items for inclusion into the test battery. The new items can be included in the score with the flick of a program option. There is no need for printing of new booklets and extensive field testing of series of booklets.

5.SUMMARY

Although, I have not supplied the final crowning answer to selection and classification research ; I have offered a system that in this new world can be exploited to process and quickly utilize any forthcoming answers to the composition of the new forces.

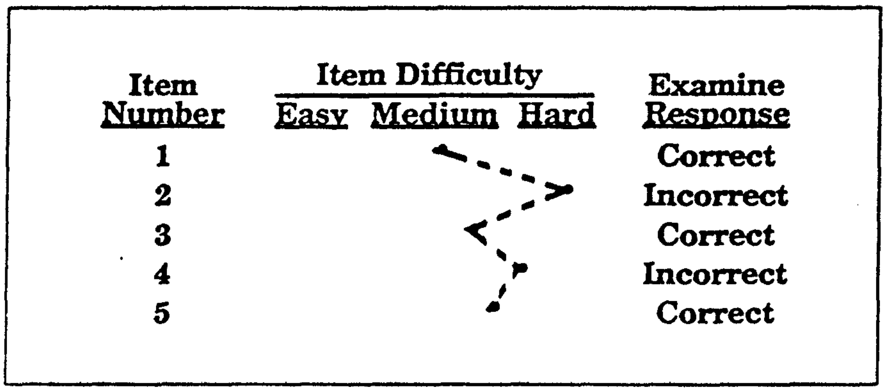


Figure 1. Illustrative Five-Item CAT

Hispanic Underrepresentation in Navy's Blue-Collar Civilian Labor Force: Organizational Perspectives

Paul Rosenfeld, Jack E. Edwards, Marie D. Thomas, & Stephanie Booth-Kewley
Navy Personnel Research & Development Center
San Diego, CA 92152-6800

Problem Statement

The U.S. Navy has had difficulty attracting Hispanics to its civilian work force despite the increasing number of Hispanics in the U.S. civilian labor force. Very little research has addressed organizational factors relevant to Hispanic underrepresentation in work settings. This paper presents findings from one phase of a Navy-wide effort to determine organizational factors related to Hispanic underrepresentation.

Background

The Hispanic component of the U.S. population increased 53% during the 1980s and currently represents 9% of the population (Stone & Castaneda, 1991). This dramatic growth will likely continue as the size and influence of the Hispanic work force continues to increase (Cattan, 1988). With high birth and immigration rates, Hispanics will surpass blacks to become the largest U.S. minority early in the twenty-first century. Despite these increases, the representation of Hispanics in the Navy's civilian work force has remained at under four percent (Edwards, M. D. Thomas, & Burch, 1992).

In recognition of the difficulties in attracting Hispanics to its civilian work force, the Navy requested that research be conducted to identify the causes of Hispanic underrepresentation and to recommend ways of overcoming it. In 1985, the Navy Personnel Research and Development Center began a multi-year Equal Employment Opportunity (EEO) Enhancement research project. Among the project's goals was the identification of cultural, individual-difference, and organizational barriers that may be preventing Hispanics from attaining employment in the civilian Navy work force.

Research conducted under the auspicies of the EEO Enhancement research project has attempted to accurately define the scope of the Hispanic under-representation problem (Edwards & P. J. Thomas, 1989; P. J. Thomas, 1987), reviewed the literature on attitudes and demographics related to work outcomes (Edwards, 1988), assessed the geographic mobility of Hispanics for employment (Edwards, Rosenfeld, P. J. Thomas, & M. D. Thomas, in press), compared the responses of Hispanic and nonHispanic new white employees on an organizational survey (Edwards, Rosenfeld, & P. J. Thomas, 1991), and looked at job turnover among nonHispanic white and Hispanic employees (Booth-Kewley, Rosenfeld, & Edwards, 1990).

There were at least four major research findings.

(1) Navy may have been inaccurately estimating its Hispanic goals and undercounting the percentage of Hispanics in its work force (P. J. Thomas, 1987).

(2) Hispanics and nonHispanic whites had generally similar responses on an organizational entry survey, although low acculturated Hispanics had a greater need for role clarity (Edwards et al., 1991).

(3) Hispanics are as likely to move for employment as nonHispanic whites and Blacks if incentives were high or if the new employment areas had high Hispanic concentrations (Edwards

et al., in press).

4) Low acculturated Hispanics had significantly higher turnover rates than high acculturated Hispanics or nonHispanic whites (Booth-Kewley, et al., 1990).

While the individual difference and cultural barriers to Hispanic representation have been dealt with in previous studies of the EEO enhancement research project (see Edwards et al., 1992, for a review), there has been very little research relating organizational factors to Hispanic underrepresentation. Furthermore, surveys that have gathered background and attitudinal data from newly hired Hispanic Navy civilian employees (Edwards et al., 1991) cannot assess the obstacles encountered when implementing EEO programs. With Hispanic representation in Navy's civilian labor force still well below many civilian and government agencies (Edwards et al., 1992), possible organizational barriers may be impeding Navy from attaining its goal of full EEO.

Method

For each of 30 Navy activities, questionnaires were sent to civilians in the following EEO-related positions: Deputy EEO Officer (DEEOO), Federal Women's Program Manager (FWPM), Hispanic Employee Program Manager (HEPM), and Civilian Personnel Officer/Industrial Relations Director (CPO). Surveys were returned by 28 DEEOOs, 28 FWPMs, 22 HEPMs, and 27 CPOs. The questionnaires contained a core of items that were identical in all four versions as well as individual items of particular relevance to the position of the individual completing it. All four questionnaires contained the following elements: respondent background information and demographics, and attitudinal items concerning EEO issues and the role of the EEO department. Furthermore, the DEEOOs and CPOs were asked a series of questions relating to reasons for underrepresentation of Hispanics and women in Navy civilian blue-collar jobs. HEPMs and FWPMs were asked to complete only those items that were related to Hispanics and women, respectively. Finally, each of the questionnaires contained additional items relevant to the particular group the respondent was representing.

Results

Demographics. Twenty-eight percent of the respondents identified themselves as Hispanics. Most HEPMs (71%) were Hispanic but relatively fewer FWPMs (25%), CPOs (12%) and DEEOOs (11%) were Hispanic. Relative to the other three groups, HEPMs had less experience in civil service, fewer years in their current job, and less time in their job series. Of the four groups, HEPMs had the lowest average government pay level. HEPMs had less years in EEO/personnel functions than FWPMs, DEEOOs, and CPOs. Interestingly, HEPMs had the highest number of self-reported EEO and personnel college courses.

EEO Perceptions. When respondents were asked to indicate which group(s) were disadvantaged, they most frequently chose women (73%), followed by Hispanics (62%), blacks (56%), handicapped individuals/disabled veterans (51%), American Indians (39%), Asians and Pacific Islanders (17%), older workers (15%), and whites (5%). While 62% of the sample mentioned Hispanics as being disadvantaged, this judgment was much more common among Hispanics (82%) than non-Hispanics (42%). The most frequently mentioned ways in which respondents felt Hispanics were disadvantaged was in simple underrepresentation in the work force (29%), and not being given enough opportunities (14%).

All respondents except for the FWPMs indicated their amount of agreement (1=strongly

disagree... 5=strongly agree) with 16 items regarding Hispanics. Across all respondents, the strongest opinion given was for the statement, "Getting a Federal job takes too long." (M=4.1). There was a slight tendency towards agreement that job advertisements do not reach Hispanics (M=3.2), Hispanics "lack technical training" (M=3.1), and "Many Hispanics are reluctant to move to a new location for a job" (M=3.2).

The opinions of the HEPMs, DEEOOs and CPOs differed for several items. For the item with the largest divergence, HEPMs generally agreed (M=3.6) that "Selecting officials pick Hispanics last", whereas DEEOOs and CPOs generally disagreed (Ms=2.6, 2.1). This difference was statistically significant, F (2, 67)= 6.62, p <.002. Marginally significant differences were obtained for "Many Hispanics have difficulty completing the SF-171 application form" (p=.06) and "Many Hispanics do not check on the status of their application after they file" (p=.08).

While most respondents (M=4.3) felt adequately trained to handle EEO issues, HEPMs (M=4.0) agreed with this statement less strongly than did the others (M=4.6). In fact, in response to the item, "The HEP manager does not have the knowledge, skills, or abilities necessary to perform that job", HEPMs (M=2.7) disagreed *less* than DEEOOs and CPOs (Ms=1.8 and 2.2). This difference approached statistical significance, F (2, 67)= 2.96, p=.06. These findings suggest that HEPMs feel less able to handle EEO issues than other respondents.

Respondents were asked to rank the effectiveness of EEO programs. Their responses are presented in Table 1.

Table 1
Responses to:"How would you rank the effectiveness of the following programs at the activity/command".
(Rankings from most to least effective)

EEO Complaints	4.1
EEO training	3.6
Federal Women's	3.4
Minority and Women Recruiting	3.1
Upward Mobility	2.7
Hispanic Employment	2.4

As can be seen, the Hispanic Employment program was seen as least effective of all the EEO programs asked about.

Conclusions

Despite the fact that Hispanic underrepresentation has been recognized by Navy policymakers as a very high priority, the individuals designated to implement the Hispanic program--the HEPMs--are of lower rank, have less experience, and feel less trained than their EEO coworkers. Given the relatively lower status of the HEPM, it is not surprising that the Hispanic Employment Program was ranked the least effective of all EEO programs. While HEPMs generally strongly endorsed several possible reasons for Hispanic underrepresentation, they may be at a disadvantage in trying to advocate or implement policies aimed at addressing

these issues given their lack of power and status within the EEO structure. If progress on the Hispanic underrepresentation issue is to be made, the authority and status of the HEPM position will need to be increased.

References

Booth-Kewley, S., Rosenfeld, P., & Edwards, J. E. *(1990, November). Turnover among Navy civilian Hispanic and Anglo blue-collar workers.* Paper presented at the 32nd annual meeting of the Military Testing Association, Orange Beach, AL.

Cattan, P. (1988). The growing presence of Hispanics in the U.S. work force. *Monthly Labor Review, 111*, 9-14.

Edwards, J. E. (1988). *Work outcomes as predicted by attitudes and demographics of Hispanics and nonHispanics: A literature review* (TN 88-23). San Diego, CA: Navy Personnel Research and Development Center.

Edwards, J. E., Rosenfeld, P., & Thomas, P.J. (1991). Hispanic and non-Hispanic white new hirees in the Navy's blue-collar civilian work force: A pilot study. *Hispanic Journal of Behavioral Sciences, 13*, 412-421.

Edwards, J. E., Rosenfeld, P., Thomas, P. J., & Thomas, M. D. (in press). Willingness to relocate for employment: A survey of Hispanics, nonHispanic whites, and blacks. *Hispanic Journal of Behavioral Sciences.*

Edwards, J. E., Thomas, M. D., & Burch, R. L. (1992). Hispanic representation in the Federal government: Lessons from the Navy's equal employment opportunity enhancement program. In S. B. Knouse, P. Rosenfeld, & A. L. Culbertson (Eds.) *Hispanics in the workplace* (pp. 231-245). Newbury Park, CA: Sage.

Edwards, J. E., & Thomas, P. J. (1989). Hispanics: When has equal employment been achieved? *Personnel Journal, 68*(6), 144, 147-149.

Stone, A. & Castaneda, C. J. (1991, June 12). Vietnamese population surges: Up 135% Census reveals. *USA Today*, 3A.

Thomas, P. J. (1987). *Hispanic underrepresentation in the Navy's civilian work force: Defining the problem* (TN 87-31). San Diego, CA: Navy Personnel Research and Development Center.

The opinions expressed herein are those of the authors. They are not official and do not represent the views of the Navy department. The authors gratefully acknowledge the assistance of Patricia J. Thomas and Luis Joseph.

Equal Opportunity Perceptions of Filipinos in the U.S. Navy

Dora Silva-Jalonen & Paul Rosenfeld
Women and Multicultural Research Office
Navy Personnel Research & Development Center

Introduction

There has been increasing interest in the area of equal opportunity (EO), in military organizations. Past military EO research has, however, focused mainly on whites, blacks, and more recently Hispanics (Rosenfeld, Culbertson, Booth-Kewley and Magnusson, 1992). In the Navy, previous studies indicated that blacks perceived less EO than whites and Hispanics; while Hispanic perceptions were slightly less positive than whites (Rosenfeld, et al., 1992). Very little consideration has been given to other racial/ethnic groups, in some measure due to their statistically lower representation in the Navy's active-duty force.

One group that has had a long association with the Navy is Filipinos, both those who are native Filipino citizens and have been allowed to enlist in the Navy and those who are Filipino Americans. The present study looks specifically at Filipino responses to the 1989 and 1991 Navy Equal Opportunity/Sexual Harassment (NEOSH) Surveys to determine how Filipino perceptions of EO compare with the white majority group. Due to space limitations, the perceptions of blacks and Hispanics are not considered.

Although Filipino-Americans serve in all branches of the military, the Navy is the only branch that has actively recruited and employed native Filipinos. As of June 30, 1992 there were a total of 19,221 Filipinos in the active-duty Navy. This represents 3.4% of the total force (3.8% enlisted; 0.9% officers) making Filipinos the fourth-largest group in the Navy after whites (72.4%), blacks (15.9%), and Hispanics (6.2%) (Naval Military Personnel Command, 1992).

As a result of the closing of Subic Bay Naval Complex, the U.S. Navy is no-longer enlisting native Filipinos. However, those who have previously enlisted are allowed to remain in the Navy. Furthermore, the continued predicted growth of the Filipino population in the U.S. (Oades, 1990), makes the EO perceptions of Filipinos an area of continuing interest.

Approach

One method of understanding the current status of Filipinos in the Navy is through perceptual data gathered on attitude surveys. Previous Navy surveys have suggested that Filipinos have more positive perceptions than whites or other minorities. An unpublished study by a Navy contractor that addressed EO issues reported that Filipinos, in 1979, had higher mean responses than whites on the Navy's Human Resource Management (HRM) survey (described in Thomas & Conway, 1983). Thomas and Conway (1983) replicated these findings based on an analysis of HRM surveys completed in 1980 and 1981. As they noted, "Filipinos...viewed the Navy's organizational climate more favorably than did whites" (Thomas & Conway, 1983, p. 9). However, these were the last Navy surveys to consider perceptions related to EO until the presently described NEOSH survey.

The purposes in administering the 1989 and 1991 NEOSH surveys were to assess the current perceptions of EO in the Navy, compare the results with future administrations, and focus on EO areas of concern to Navy policymakers. Of special interest to the present study,

were the EO perceptions of Filipinos in the Navy.

NEOSH 1989

Method

A total of 10,070 surveys were mailed in October 1989 directly to respondents randomly selected from active-duty Navy personnel. Because of the sensitive nature of the topic, the survey was to be returned anonymously. From the surveys that were delivered, 5,558 completed surveys were received, a corrected response rate of 60% (Rosenfeld, et al., 1992).

Subjects

One percent of the respondents indicated that they were Filipino (*N*=51). Due to the small number of Filipino Officers (*N*=8), these analyses focused only on enlisted male Filipinos (*N*=43). Filipino women were also not included in these analyses due to the low number of Filipino women represented both in the officer (*N*=1) and enlisted (*N*=9) samples. There were 600 white enlisted male respondents.

1989 NEOSH Survey

The 1989 NEOSH survey consisted of 86 items having to do with EO and sexual harassment as well as demographic questions. The present focus is on the 65 EO items. The items relating to EO were combined into modules based on item content, item-response intercorrelations and factor analyses. Table 1 lists the modules and their reliabililities. Negatively worded items were reverse-scored so that for all modules, a high score indicates a more positive response and a low score indicates a more negative response.

Table 1

Module Reliabilites for NEOSH 1989 Survey Modules
Enlisted Personnel

Module	Reliability
Assignments	.52
Training	.76
Leadership	.76
Communications	.82
Interpersonal Relations	.88
Grievances	.80
Discipline	.67
Performance Evaluation	.76
Navy Satisfaction	.76

Results

A total EO score was calculated by averaging the individual modules into one overall module score. An analysis of variance (ANOVA) comparing white and Filipino enlisted males found that there were no significant difference between the way white enlisted males (M=3.70) and Filipino enlisted males (M=3.56) perceived EO issues, $F(1,591) = 1.66$, $p > .20$.

Although the overall module scores were not significantly different, there were some differences at the individual module level. White enlisted males were significantly more positive than their Filipino counterparts for the Assignments, Interpersonal Relations, and Discipline modules (*p*'s <.05). Interestingly for most of the other modules, enlisted Filipino Males had more positive perceptions than the white enlisted Males, however the differences were not statistically significant (*p*'s >.05). These results are presented in Table 2.

Table 2

Comparison of 1989 NEOSH Module Scores
White Enlisted Males vs. Filipino Enlisted Males.

Module Subscales	White *M*	Filipino *M*	*F*
Assignments	3.65	3.38	5.68*
Training	3.66	3.74	.19 NS
Leadership	4.01	3.99	.03 NS
Communications	3.73	3.76	.05 NS
Interpersonal Relations	3.78	3.40	6.83**
Grievances	3.51	3.61	.35 NS
Discipline	3.90	3.56	6.82**
Performance Evaluation	3.32	3.30	.01 NS
Navy Satisfaction	3.51	3.55	.03 NS
Total(Combined Modules)	**3.70**	**3.56**	**1.66 NS**

*p<.05, **p <.01

NEOSH 1991

Method

NEOSH surveys were directly mailed to active-duty Navy personnel (N=12,006) in October 1991. Of the surveys that were delivered, a total of 5,225 completed surveys were received, resulting in a corrected response rate of 48%

Subjects

As in 1989, about 1% of the respondents indicated that they were Filipino. These analyses focused once again solely on enlisted Filipino males (*N*=25) and enlisted white males (*N*=389).

1991 NEOSH Survey

The NEOSH survey consisted of a total of 147 items covering the areas of EO, fraternization, sexual harrassment, rape and sexual assault. The present focus is on the 62 EO items. In 1991, a module was added for assessing Promotions/Advancement, resulting in a total of eleven EO modules. Table 3 lists the modules and their reliabilities.

Table 3
Module Reliabillities for NEOSH 1991 Survey Modules
Enlisted Personnel

Module	Reliability
Assignments	.82
Training	.78
Leadership	.81
Communications	.85
Interpersonal Relations	.76
Grievances	.82
Discipline	.85
Performance Evaluation	.68
Promotions/Advancement	.66
Social Support	.67
General Issues	.86

Results

A total EO score was calculated by averaging the individual modules into one overall module score. An ANOVA comparing White and Filipino enlisted males found that Filipinos were more positive (M=3.94) than whites (M=3.46), $F(1,129) = 7.00$ ($p < .01$) on the EO portion of the NEOSH.

At the individual module level, ANOVA results indicated that Filipino enlisted males were significantly more positive than white enlisted males on almost all of the modules. The Interpersonal Relations module and the Disicipline module both failed to reach significance ($p > .05$). These results are summarized in Table 4.

Table 4
Comparison of 1991 NEOSH Module Scores
White Enlisted Males vs. Filipino Enlisted Males.

Module Subscales	White M	Filipino M	F
Assignments	3.41	3.90	6.94**
Training	3.47	4.04	8.99***
Leadership	3.75	4.19	7.21**
Communications	3.60	4.13	8.51***
Interpersonal Relations	3.80	3.94	.60 NS
Grievances	3.44	3.90	6.90**
Discipline	3.91	3.92	.01 NS
Performance Evaluation	3.65	4.15	9.23***
Promotions Advancement	3.27	3.65	5.68*
Social Support	3.18	3.71	8.98***
Navy Satisfaction	3.25	3.93	10.13**
Total (Combined Modules)	**3.46**	**3.94**	**7.00****

*$p<.05$ **$p<.01$ ***$p<.001$

Conclusions

Only partial support was found for the notion, based on previous findings (e.g., Thomas & Conway, 1983), that Filipinos would have more positive EO perceptions than whites. While this was generally true for the 1991 data, the 1989 NEOSH found that the groups did not differ on overall EO perceptions and whites were more positive for several of the individual modules. It is unclear why the results were stronger in 1991 than in 1989, and given the small Filipino sample sizes caution should be used in interpreting these findings. Future administrations of the biennial NEOSH survey may shed further light on this issue.

Given that over 3% of the Navy is Filipino, it is surprising that so little work has looked at this group. Unlike other minority groups, Filipinos in the Navy are composed of two distinct groups--those who have been recruited directly from the Philippines and those who are Americans of Filipino origin. One shortcoming of the present study is that it was not possible to distinguish which of the two Filipino groups the respondents belonged to. Indeed the Navy does not distinguish between native Filipinos and Filipino Americans in its demographic breakdowns (Naval Military Personnel Command, 1992). Future research should compare and contrast these two Filipino subgroups.

Both the 1989 and 1991 results indicate that the Interpersonal Relations and Discipline modules stand out from the others. In 1989, Filipinos were significantly less positive on these modules than whites. In 1991, these were the only two modules where Filipinos were not significantly more positive. This is similar to other findings on the NEOSH (Rosenfeld, et al, 1992) where the largest differences between whites and minority members is on items related to interpersonal relations (i.e., items about discrimination) and discipline. Because items within these modules ask about perceptions of discrimination and fairness in punishment they assess the negative aspects of EO climate. They thus may be most salient to non-white Navy personnel.

In sum, the results of the 1989 and 1991 NEOSH surveys provide mixed evidence that Filipinos have more positive EO perceptions than whites. Future administrations of the NEOSH survey should further clarify the nature of Filipino EO perceptions.

References

Naval Military Personnel Command (1992, September). *Navy-wide demographic data for third quarter FY-92*. Washington, DC: Department of the Navy.

Oades, R., (1990, June 28). The Browning of San Diego: The Brown skins you see are not those of sun-tanned white San Diegans but those of the growning number of Filipinos. *San Diego Asian Journal*, pp. 5, 6.

Rosenfeld, P. Culbertson, A.L., Booth-Kewley, S. and Magnusson, P. (1992, May). *Assessment of Equal Opportunity Climate: Results of the 1989 Navy-wide Survey* (NPRDC-TR-92-14). San Diego: Navy Personnel Research and Development Center.

Thomas, P.J. & Conway, S. (1983, January). *Racial/Ethnic and gender difference in responses to the human resource management survey of personnel assigned to the Atlantic and Pacific fleets*. (NPRDC-SR-83-10). San Diego: Navy Personnel Research and Development Center.

The opinions expressed in this paper are those of the authors. They are not official and do not represent the views of the Navy department.

SUPERVISOR'S GENDER AND RACE AFFECT NAVY BLACK FEMALES' EQUAL OPPORTUNITY PERCEPTIONS[1]

Carol E. Newell, Paul Rosenfeld, & Amy L. Culbertson
Women & Multicultural Research Office
Navy Personnel Research & Development Center

Black Females in the Military

Current reports reveal that black women comprise 35% of all women in the Navy (U.S. General Accounting Office, 1991). Although represented in such a large percentage, very little research has focused on this group. Brenda Moore (1991a) reviewed the participation rates of black females in the U.S. military from past to present. Her findings were that the percentage of black females in the armed forces has increased from one-half percent to almost four percent during the years 1974 through 1989. Like their male counterparts, most black females joined the Army. In addition, black females served longer terms in the services than Hispanic and white females, were more likely than Hispanic and white females to be enlisted than officer personnel, and were more likely to receive less technical training. In another paper, Moore (1991b) reflected upon the status and trends of black females in the Navy. Similar to the results of the earlier study, Moore found that there was an overrepresentation of enlisted black females and an underrepresentation of black female officers when compared to the population. In terms of education, black and white females had equal levels of attainment, but black females scored lower on military aptitude tests than white females. While Moore's effort to learn more about Navy life for the black female focused on demographic and career issues another approach is to consider perceptual data regarding the attitudes of black Navy women.

The 1989 NEOSH

In order to assess the perception of equal opportunity and the occurrence of sexual harassment, the 1989 Navy Equal Opportunity/Sexual Harassment (NEOSH) Survey was mailed to a random sample of Navy personnel. Due to the nature of the survey, blacks, Hispanics, and females were oversampled in order to insure that they were adequately represented. Of the 9,309 surveys that were delivered, 5,558 were analyzed, resulting in a corrected response rate of 60%. The NEOSH survey consisted of 65 questions on Equal Opportunity (EO) which were in Likert format, with a five point scale ranging from strongly agree to strongly disagree (Rosenfeld, Culbertson, Booth-Kewley, & Magnusson, 1992). There were nine modules or general areas on the EO section, such as Assignments, Training, and Promotions/Advancement. Internal consistency reliabilities were determined for each module, and they ranged from .52 to .88 for the enlisted sample and from .62 to .87 for the officers (Table 1). The remaining questions surveyed sexual harassment issues (cf. Culbertson, Rosenfeld, Booth-Kewley, & Magnusson, 1992).

Table 1. Module Reliabilities for NEOSH Survey Modules

	Reliability	
Module	Enlisted	Officer
Assignments	.52	.62
Training	.76	.82
Leadership	.76	.74

1The opinions expressed herein are those of the authors. They are not official and do not represent the views of the Navy department.

Module	Enlisted	Officer
Communications	.82	.80
Interpersonal Relations	.88	.87
Grievances	.80	.80
Discipline	.67	.76
Performance Evaluation	.76	.74
Navy Satisfaction	.76	.74

Subjects were categorized according to their gender and race resulting in six groups for officers and enlisted: Hispanic males, Hispanic females, white males, white females, black males and black females. Table 2 shows the overall EO module means of each of these groups, which were calculated by combining all questions from each module. Overall results of the survey indicated that males were more satisfied than females, officers were more satisfied than enlisted personnel, and whites were more satisfied than Hispanics, who were more satisfied than blacks. Furthermore, enlisted and officer black females are the least satisfied group with regard to EO topics. This trend was also evident for each of the individual module means.

In order to determine why the perceptions of black Navy women were least positive of any group a further set of analyses was conducted. These analyses focused on the race and gender of black women's supervisors. The purpose of the present study is to determine if certain factors, such as the race and gender of these black females' supervisors will reduce these negative perceptions. Research in the area of mentoring suggest that this may be true.

Table 2: Overall Module Means for Officer and Enlisted Respondents by Gender and Race of Respondent

	WM	WF	HM	HF	BM	BF
Enlisted Overall Module Mean	3.70	3.53	3.50	3.39	3.32	3.17
Officer Overall Module Mean	4.16	3.79	3.94	3.68	3.66	3.43

Research on Mentoring Minorities and Women

Literature on mentoring racial/ethnic groups and women has found that it is most effective when the mentor and protegee are of the same race or same gender (Knouse, 1992; Thomas, 1990). One reason for this is that in same race mentoring relationships both members identify with each other because of their similar cultural background. This identification leads to better rapport between them, which may be lacking in different race mentoring relationships (Knouse, 1992). Additionally, same race mentoring relationships are more effective because these mentors are able to recognize and empathize with special problems that protegees may encounter on the job as a result of their minority status, and also offer guidance on how to effectively resolve these problems based on their previous experience (Knouse, 1992). Moreover, minority mentors can serve as a role model of appropriate behaviors and values in the organization that the minority protegee may be unaware of because they differ from the behaviors and values of the protegee' s culture (Knouse, 1992). As regards gender, Thomas (1990) and Gaskill (1991) found that same gender mentoring relationships offer more career and psychosocial support for protegees than different gender mentoring relationships.

Recent research has referred to black females and other minority females as "double minorities", disadvantaged as a consequence of both gender and race. Studies have shown that these effects are additive (McNett, Taylor, & Scott, 1985), and the findings of the 1989 NEOSH survey lend support to this. Black females were less satisfied than both black males and white females. As mentioned, research in the area of mentoring has found that mentoring minorities and females works best when the mentor and protegee are of the same race or gender. This research suggests that if personnel have a supervisor that they can identify with on both variables, race and gender, then their perceptions will be more positive than personnel with supervisors of a different race and gender. In this study, it is hypothesized that black females

who had same sex and same race supervisors were more positive in their responses to NEOSH than black females with other race/gender supervisors.

Method

Subjects:

Due to the small number of black female officers who had black supervisors, the data analysis focused on the black enlisted females ($N = 523$).

Procedure:

The data from the EO section of the 1989 NEOSH survey were re-analyzed. Two demographic questions, 1) Are you and your immediate supervisor members of the same race/ethnic group and 2) Are you and your immediate supervisor the same sex, were used to categorize the black enlisted females into groups according to their supervisor type: non black male (N=367), non black female (N=87), black male (N=47), and black female (N=22). Analyses of Variance (ANOVAs) were performed in order to determine if the four groups differed on where they were homeported, their paygrade, and their number of years in service. Subsequently, a two by two design was utilized, with sex of supervisor and race of supervisor serving as the independent variables and the module means serving as the dependent variables. ANOVAs were performed on each EO module. Additional ANOVAs were performed on individual questions within certain modules.

Results

ANOVAs revealed that there were no significant differences between the groups on where they were homeported, $F(3, 523) = .31$, $p > .10$; their paygrade, $F(3,523) = .99$, $p > .10$; or their years in service, $F(3,523) = .30$, $p > .10$. On average, the black enlisted females who completed the survey were homeported on shore, had served in the Navy for 0-4 years, and were in the E-4 through E-6 paygrades.

Results showed that the interaction between race and sex of supervisor was significant on two modules, Leadership, $F(1, 523) = 5.216$, $p < .05$, and Grievances $F(1, 523) = 5.198$, $p < .05$. On both of these modules ANOVAs indicated that subjects with black female supervisors were significantly more positive than the other groups. Scheffe tests were performed on both of these modules. For the Leadership module there was a significant difference between subjects with black female supervisors and those with non black female supervisors $F(1, 523) = 4.042$, $p = .01$. The Grievances module approached a significant difference $F(1,523) = 2.443$, $p = .06$) between subjects with black female supervisors and those with non black female supervisors.

There were no significant differences on the remaining seven modules. However, subjects with black female supervisors had the highest module means on each module except Training (Table 3). On the Training module, subjects with black female supervisors had the lowest module mean score. An ANOVA was done on the overall module mean. Although there was not a significant interaction between the gender and race of the supervisor, the result was that subjects with black female supervisors had the highest mean, indicating that they were the most positive of the four groups (Table 3).

Table 3. Results of ANOVA: Means of Interactions Between Race & Gender of Supervisor

	Supervisor Type			
Module	BF	BM	NBM	NBF
Assignments	3.14	3.14	3.03	3.08
Training	3.05	3.44	3.11	3.14
Leadership	3.92[a]	3.55	3.43	3.27[a]**

Module	BF	BM	NBM	NBF
Communications	3.73	3.39	3.35	3.48
Interpersonal Relations	3.47	3.29	3.21	3.18
Grievances	3.51	3.06	3.09	2.88**
Discipline	3.15	2.94	2.98	2.79
Performance Evaluation	3.33	2.86	2.88	2.98
Navy Satisfaction	3.45	3.27	3.22	3.16
Overall Mean	3.49	3.20	3.15	3.15

** = Significant Difference ($p < .05$).

[a] = Means with same superscript are significantly different ($p < .05$).

Additional ANOVAs were performed on individual questions in the Leadership and Grievances Modules. For each of these questions, subjects with black female supervisors were significantly more positive than subjects with other supervisors (Table 4).

Table 4. Results from ANOVA: Means of Interaction Between Race & Gender of Supervisor on Questions on Leadership and Grievances Modules

Question from Leadership Module	BF	Supervisor Type BM	NBM	NBF
CO Aware of Discrimination and Sexual Harassment That May Occur at This Command	4.00	3.08	3.31	3.35**
Questions from Grievances Module				
Chain of Command is an Effective Way to Resolve EO Problems	3.41	2.74	3.04	2.83**
I Would Talk to my Supervisor if I Felt Discriminated Against While at Work	4.10	3.78	3.53	3.52**
Filing a Grievance Would not Hurt my Navy Career	3.18	2.37	2.81	2.70**

** = Significant Difference ($p < .05$).

Discussion

The results of the present study were that for two of the nine EO modules, Leadership and Grievances, black female employees with same race/gender supervisors were significantly more positive than employees with other race/gender supervisors. The remaining seven modules did not reach significance. However, for six of these modules, employees with black female supervisors had the highest means. For the Training module employees with black female supervisors had the lowest mean.

When one examines the means of most of the EO modules, subjects with black female supervisors had the highest means. This trend was the same for the overall EO module. On the Leadership module, a significant difference between subjects with black female supervisors and those with non black female supervisors was obtained. Although subjects had supervisors with whom they could identify on one dimension, their gender, these subjects were still the least satisfied with regards to Leadership of the four groups. In addition, on five of the nine EO modules subjects with non black female supervisors had the lowest means. These results suggest that both gender and race are important variables to consider when assessing black females' perceptions. Therefore, the hypothesis that black females with same gender and same race supervisors would be significantly more positive than black females with other race/gender supervisors obtained some support.

One of the weaknesses of this study is that it was not the main focus of the original study, but a post hoc analysis. Another limitation is tne small number of subjects with black supervisors, especially black females. A larger number of subjects in the black supervisor categories (i.e., female and male) would have been beneficial. Also, the hypothesis could not be tested on other minority females (i.e. Hispanic, Filipino, etc.) because of the small number who had same gender and same race supervisors.

Future analysis should be done with a larger sample of black supervisors However, this will be difficult because of the small number of minority females who have same gender and/or race supervisors in any organization. In addition, future research on comparisons between minority females of different racial backgrounds on these issues is warranted in order to determine if there are differences between races. Another area that should be investigated is research into why minority racial groups might perceive EO differently. Also, while black males' perceptions were not the focus of this paper, they should be examined in the future.

References

Culbertson, A.L., Rosenfeld, P., Booth-Kewley, S., & Magnusson, P. (1992). Assessment of Sexual Harassment in the Navy: Results of the 1989 Navy-wide Survey (Technical Report No. 92-11). San Diego, CA: Navy Personnel Research and Development Center.

Defense Equal Opportunity Management Institute. (1991, September). Semi-annual race/ethnic/gender profile of the Department of Defense active forces, reserve forces, & the U.S. Coast Guard. Patrick AFB, FL: Defense Equal Opportunity Management Institute, Research Division.

Gaskill, L.R. (1991). Same Sex and Cross Sex Mentoring of Female protegees: A Comparative Analysis. Career Development Quarterly, 40, 48-63.

Knouse, S.B. (1992). The Mentoring Process for Hispanics. In Knouse, S.B., Rosenfeld, P., & Culbertson, A.L. (Eds.). Hispanics in the Workplace (pp. 137-150). Newbury Park: Sage Publications.

McNett, I., Taylor, L., & Scott, L. (1985). Minority Women: Doubly Disadvantaged. In Alice G. Sargent (Ed.). Beyond Sex Roles, (pp. 226-232). St. Paul Minnesota: West Publishing Co.

Moore, B.L. (1991a). African-American Women in the U.S. Military. Armed Forces & Society, 17, 363-384.

Moore, B.L. (1991b). Issues of Black Women in the United States Navy: Preliminary Findings. (Unpublished letter report). San Diego: Navy Personnel Research and Development Center.

Rosenfeld, P., Culbertson, A.L., Booth-Kewley, S., & Magnusson, P. (1992). Assessment of Equal Opportunity Climate: Results of the 1989 Navy-wide Survey (Technical Report No. 92-14). San Diego, CA: Navy Personnel Research and Development Center.

Scott, N.E. (1989). Differences in Mentor Relationships of Non-White and White Female Professionals and Organizational Mobility: A Review of the Literature. Psychology, A Journal of Human Behavior, 26, 23-26.

Thomas, D.A. (1990). The Impact of Race on Managers' Experiences of Developmental Relationships (Mentoring and Sponsorship): An Intra-Organizational Study. Journal of Organizational Behavior, 11, 479-492.

U.S. General Accounting Office (1991, February). Military Personnel: Composition of the active-duty forces by race or national origin identification and by gender (GAO/NSIAD - 91-134FS). Gaithersburg, MD: United States General Accounting Office.

Author Index

A

Alderks, C. E. 420, 426
Alderton, D. L. 33, 461
Allender, L. 804
Allen, J. P. 245
Arabian, J. M. 792
Armitage, M. A. 164, 170

B

Baker, H. G. 158, 888
Barnes, J. D. 396
Bartone, P. T. 257
Bennett, W. R., Jr. 724
Besetsny, L. K. 119
Bierbaum, C. 804
Bloxom, B. 16, 455
Booth-Kewley, S. 182, 269, 929
Borman, W. C. 402
Bosshardt, M. J. 636
Brady, E. J. 553
Brannan, T. S. 610
Bucher, E. W. 671
Buck, L. S. 763
Busciglio, H. H. 564, 570
Bush, B. J. 286

C

Caldwell, J. A. 514
Campbell, J. P. 676, 682, 688, 694
Carreta, T. R. 837
Carroll, E. R. 194
Christ, R. E. 290
Coccia, F. 831
Cooke, R. S. 373
Corts, D. B. 170
Costello, G. S. 205
Crawford, K. S. 636
Crowley, J. S. 514
Culbertson, A. L. 938
Curran, L. T. 467, 479

D

Datko, L. M. 449
Devlin, S. E. 146
Dillon, R. F. 526, 600, 605, 610, 616
Dittmar, M. J. 827
Dolgin, D. L. 251
Doyle, K. A. 768
Dorin, D. E. 113
Doucette, R. E. 509
Driskill, W. E. 827, 849
Dumestre, J. C. 861
Dunlap, W. P. 408
Dutcher, J. S. 373, 379, 438

E

Earles, J. A. 559
Ebenrett, H. J. 654
Edwards, J. E. 929

F

Fertig, J. B. 245
Fischer, S. G. 200
Flemming, S. B. 532
Fogel, L. J. 493
Folchi, J. S. 146
Ford, J. K. 281
Freville, M. E. 538
Fullerton, G. F. 497

G

Gettman, D. J. 51
Gibson, E. G. 855
Gifford, R. K. 257
Glick, G. 712
Goehring, D. J. 211
Goldman, L. A. 757
Goodwin, G. F. 384
Graham, S. E. 217
Greer, R. J. 600

Gregory, D. J. 164
Griffin, G. R. 861
Guzaitis, J. G. 712

H

Hand, D. K. 729
Hansen, D. 205
Harman, J. 95
Harris, B. C. 443, 594
Harris, C. A. 605
Harris, D. A. 396
Harris, J. A. 700
Harris, J. C. 745
Harville, D. L. 821
Hay, M. S. 107
Heller, C. B. 438
Helton, K. T. 251
Hetter, R. D. 16, 45
Hogan, P. F. 473
Holz, R. F. 798
Hufcutt, A. I. 724
Huguley, H., III 816
Humphreys, L. 900
Hunt, D. P. 497

I

Idar, I. 361, 786

K

Kennedy, R. S. 408
Kieckhaefer, W. F. 620
Kilcullen, R. N. 194
Kinney, J. C. 152
Klein, G. 487
Knapp, D. J. 676, 688, 694
Knight, J. R. 831
Kokorian, A. 306

L

Lakhani, H. 547
Larson, G. E. 33
Laurence, J. H. 128
Lescreve, F. 310
Lesher, L. L. 296
Lewis, C. 331
Lickteig, C. W. 588
Lindsay, T. J. 100
Lockhart, J. M. 223, 228

M

Main, R. G. 54
Maisano, R. E. 420, 426
Marlowe, D. H. 257
Martell, K. A. 792
Martin, C. J. 316
Martin, J. A. 239
Matherne, R. J. 74
Matyuf, M. M. 402
Mayhill, T. K. 745
McBride, J. R. 125, 467, 473
McAnulty, D. M. 804
McCauley, D. E., Jr. 176
McCloy, R. A. 396
McDonald, B. 361
McWhite, P. 344
Melter, A. H. 665
Mitchell, J. L. 729, 831, 849
Miller, M. R. 821
Miller, T. E. 487
Modrick, J. A. 751
Moore, D. L. 780
Morales, M. M. 281
Moreno, K. E. 22, 45
Moskow, D. B. 872
Muraida, D. J. 355

N

Neal, G. L. 810
Newell, C. E. 938
Nogami, G. Y. 89

O

Oppler, S. H. 188

P

Palmer, D. R. 564
Paniesin, R. 296
Paquette, L. J. 509
Paquette, S. P. 233
Park, R. K. 164, 170
Parker, J. P. 630, 642
Perez, R. S. 217
Peterson, N. G. 894, 900
Phalen, W. J. 729
Phillips, D. C. 390
Power, D. A. 200
Price, J. S. 741

Q

Quenette, M. A. 877

R

Ree, M. J. 559, 837, 912
Reeder, J. M. 275
Reynolds, D. H. 682, 694
Rice, V. 263
Riedel, J. A. 642
Rodel, G. W. 660
Rodgers, W. 867
Rosenfeld, P. 929, 933, 938
Rosenthal, M. B. 379
Rosse, R. L. 894, 900
Ruck, H. W. 724, 849
Rumsey, M. G. 918
Russell, T. L. 676, 682, 694, 894, 900

S

Salter, C. A. 514
Sands, W. A. 12, 837
Scaramozzino, J. A. 786
Scarville, J. 443
Schuette, D. W. 745
Segall, D. O. 16, 22, 27
Sego, D. 281
Sellman, W. S. 3, 125, 336, 485
Sharon, A. T. 275
Sharp, M. 263
Sheposh, J. P. 373, 379, 438
Shrum, R. C. 831
Siebold, G. L. 100
Silva-Jalonen, D. E. 933
Silver, J. D. 223, 228
Simpson, H. 53
Skinner, J. 821
Smart, D. L. 588
Smith, R. L. 514
Spector, J. M. 74, 355
Stanley, P. P., II 745
Steege, F. W. 648
Steinberg, A. G. 443
Stephenson, J. A. 774
Stephenson, S. D. 367, 414, 774
Stone, L. A. 576
Stouffer, J. M. 582
St-Pierre, R. 205
Street, D. R. 251
Streufert, S. 349

T

Tatsuoka, K. 328
Teachout, M. S. 281
Thain, J. W. 906
Tharion, W. 263
Thayer, D. 322
Thomas, M. D. 182, 929
Thor, K. K. 194
Tremble, T. R., Jr. 107, 384
Trent, T. 146
Turnage, J. J. 408

V

Vaitkus, M. A. 239
Van Raay, P. B. 301
Vandivier, P. L. 432, 520
Vandivier, S. 432
Vaughan, D. S. 735, 849
Vicino, F. L. 340, 923

W

Waldkoetter, R. O. 432, 520
Ward, J. H., Jr. 849
Ward, J. L., II 296
Watson, T. W. 119
Webb, R. 616
Weissmuller, J. J. 827
Welsh, J. R. 718
White, L. A. 140, 188, 402
Wilbur, E. R. 45
Wilcove, G. L. 883
Williams, S. G. 588
Wilson, A. S. 72
Wingersky, M. 322
Wise, L. L. 455, 706
Wiskoff, M. F. 624, 630
Wolfe, J. H. 39
Wood, S. 624
Wright, K. M. 257

Y

Yadrick, R. M. 735
Young, M. C. 140, 188

Z

Ziebell, T. S. 503
Zwick, R. 322